Contents

Viewpoints

Case-Initiated COVID-19 Contact Tracing Using Anonymous Notifications (e20369)
Weibin Cheng, Chun Hao ................................................................. 7

A Pantheoretical Framework to Optimize Adherence to Healthy Lifestyle Behaviors and Medication
Adherence: The Use of Personalized Approaches to Overcome Barriers and Optimize Facilitators to Achieve
Adherence (e16429)
Azizi Seixas, Colleen Connors, Alicia Chung, Tiffany Donley, Girardin Jean-Louis .................................................. 11

Medical Device Apps: An Introduction to Regulatory Affairs for Developers (e17567)
Lina Keutzer, Ulrika Simonsson .......................................................... 23

A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over
Network and GPS (e19457)
Shaoxiong Wang, Shuizi Ding, Li Xiong ................................................. 738

Reviews

Mobile Health for Pediatric Weight Management: Systematic Scoping Review (e16214)
Louise Tully, Amanda Burlis, Jan Sorensen, Riyad El-Moslemany, Grace O’Malley .......................................................... 30

Hospital Bring-Your-Own-Device Security Challenges and Solutions: Systematic Review of Gray Literature
(e18175)
Tafheem Wani, Antonette Mendoza, Kathleen Gray .................................................. 463

The Quality of Mobile Apps Used for the Identification of Pressure Ulcers in Adults: Systematic Survey and
Review of Apps in App Stores (e14266)
Janine Koepp, Miriam Baron, Paulo Hernandes Martins, Cristine Brandenburg, Ariane Kira, Vanessa Trindade, Luis Ley Dominguez, Marcelo
Carneiro, Rejane Frozza, Lia Possuelo, Marcus De Mello Pinto, Liane Mahlmann Kipper, Bartira Pinheiro da Costa ...................... 503

Telemedicine in Chronic Wound Management: Systematic Review And Meta-Analysis (e15574)
Lihong Chen, Lihui Cheng, Wei Gao, Dawei Chen, Chun Wang, Xingwu Ran .......................................................... 577
Original Papers

Overview of Policies, Guidelines, and Standards for Active Assisted Living Data Exchange: Thematic Analysis (e15923)
Laura Fadrique, Dia Rahman, Hélène Vaillancourt, Paul Boissonneault, Tania Donovska, Plinio Morita ............................................................. 45

Perceptions of Home Telemonitoring Use Among Patients With Chronic Obstructive Pulmonary Disease: Qualitative Study (e16343)
Sara Lundell, Mari Modig, Åsa Holmner, Karin Wadell. ................................................................. 56

Evaluating Online Consumer Medication Information Systems: Comparative Online Usability Study (e16648)
Stefan Sigle, Pilar Barriga, Francisco Correa Fernández, Christian Juhra, Steffen Härtel, Christian Fegeler. .......................................................... 69

Factors Influencing Acceptance of Personal Health Record Apps for Workplace Health Promotion: Cross-Sectional Questionnaire Study (e16723)
Hyun Park, Kwang Kim, Jae Soh, Young Hyun, Sae Jang, Sol Lee, Ga Hwang, Hwa Kim. ................................. 82

Experiences of African American Breast Cancer Survivors Using Digital Scales and Activity Trackers in a Weight Gain Prevention Intervention: Qualitative Study (e16059)
Julianne Power, Deborah Tate, Carmina Valle. ................................................................. 106

Patients' Measurement Priorities for Remote Measurement Technologies to Aid Chronic Health Conditions: Qualitative Analysis (e15086)
Sara Simblett, Faith Matcham, Hannah Curtis, Ben Greer, Ashley Polhemus, Jan Novák, Jose Ferrao, Peter Gamble, Matthew Hotopf, Vaibhav Narayan, Tim Wykes, Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) Consortium ........................................... 116

Promoting Antenatal Care Attendance Through a Text Messaging Intervention in Samoa: Quasi-Experimental Study (e15890)
Jessica Watterson, Diego Castaneda, Caricia Catalani. ................................................................. 124

Smartphone-Guided Algorithms for Use by Community Volunteers to Screen and Refer People With Eye Problems in Trans Nzoia County, Kenya: Development and Validation Study (e16345)
Hillary Rono, Andrew Bastawrous, David Macleod, Cosmas Bunywera, Ronald Mambolleo, Emmanuel Wanjala, Matthew Burton. ................................. 134

The Clinical and Cost-Effectiveness of Telerehabilitation for People With Nonspecific Chronic Low Back Pain: Randomized Controlled Trial (e15375)
Francis Fatoye, Tadesse Gebrye, Clara Fatoye, Chidozie Mbacha, Mistura Olaoye, Adesola Odole, Olumide Dada. ................................................................. 148

mHealth App Prescription in Australian General Practice: Pre-Post Study (e16497)
Oyungerel Byambasuren, Elaine Beller, Tammy Hoffmann, Paul Glasziou. ................................................................. 158

Assessing the Impact of Patient-Facing Mobile Health Technology on Patient Outcomes: Retrospective Observational Cohort Study (e19333)
Courtney Bruce, Patricia Harrison, Tariq Nisar, Charlie Giammattei, Neema Tan, Caitlin Bliven, Jamie Shallcross, Aroub Khleif, Nhan Tran, Sayali Kelkar, Noreen Tobias, Ana Chavez, Dana Rivera, Angela Leong, Angela Romano, S Desai, Josh Sol, Kayla Gutierrez, Christopher Rappel, Eric Haas, Feibi Zheng, Kwan Park, Stephen Jones, Paul Barach, Roberta Schwartz. ................................................................. 171

Comparing a Social and Communication App, Telephone Intervention, and Usual Care for Diabetes Self-Management: 3-Arm Quasiexperimental Evaluation Study (e14024)
Ching-Ju Chiu, Yung-Chen Yu, Ye-Fong Du, Yi-Ching Yang, Jou-Yin Chen, Li-Ping Wong, Chanuantong Tanasugarn. ................................................................. 184
Digital Phenotyping Self-Monitoring Behaviors for Individuals With Type 2 Diabetes Mellitus: Observational Study Using Latent Class Growth Analysis (e17730)
Qing Yang, Daniel Hatch, Matthew Crowley, Allison Lewinski, Jacqueline Vaughn, Dori Steinberg, Allison Vorderstrasse, Meilin Jiang, Ryan Shaw. ................................................................. 197

User Retention and Engagement With a Mobile App Intervention to Support Self-Management in Australians With Type 1 or Type 2 Diabetes (My Care Hub): Mixed Methods Study (e17802)
Mary Adu, Usman Malabu, Aduli Malau-Aduli, Aaron Drovandi, Bunmi Malau-Aduli. ................................................................. 207

Using an Interactive App for Symptom Reporting and Management Following Pancreatic Cancer Surgery to Facilitate Person-Centered Care: Descriptive Study (e17855)
Tina Gustavell, Kay Sundberg, Ann Langius-Eklöf. ................................................................. 224

Enhancement of Neurocognitive Assessments Using Smartphone Capabilities: Systematic Review (e15517)
John Templeton, Christian Poellabauer, Sandra Schneider. ................................................................. 238

Feasibility and Utility of mHealth for the Remote Monitoring of Parkinson Disease: Ancillary Study of the PD_manager Randomized Controlled Trial (e16414)
Dimitris Gatsios, Angelo Antonini, Giovanni Gentile, Andrea Marcante, Celia Pellicano, Lucia Macchiusi, Francesca Assogna, Gianfranco Spalletta, Heather Gage, Morro Touray, Lada Timotijevic, Charo Hodgkins, Maria Chondrogiorgi, George Rigas, Dimitrios Fotiadis, Spyridon Konitsiotis. 2 5 2 5

A Web-Based Mobile App (INTERACCT App) for Adolescents Undergoing Cancer and Hematopoietic Stem Cell Transplantation Aftercare to Improve the Quality of Medical Information for Clinicians: Observational Study (e18781)
Anita Lawitschka, Stephanie Buehrer, Dorothea Bauer, Konrad Peters, Marisa Silbernagl, Natalia Zubarovskaya, Barbara Brunmair, Fares Kayali, Helmut Hlavacs, Ruth Mateus-Berr, David Riedl, Gerhard Rumpold, Christina Peters. ................................................................. 268

Feasibility and Acceptability of a Mobile Mindfulness Meditation Intervention Among Women: Intervention Study (e15943)
Ariane Rung, Evrim Oral, Lara Berghammer, Edward Peters. ................................................................. 283

Effectiveness of a 3-Month Mobile Phone–Based Behavior Change Program on Active Transportation and Physical Activity in Adults: Randomized Controlled Trial (e18531)
Anna Ek, Christina Alexandrou, Emmie Söderström, Patrick Bergman, Christina Delisle Nyström, Artur Direito, Ulf Eriksson, Pontus Henriksson, Ralph Maddison, Viva Tolle Lagerros, Marcus Bendtsen, Marie Löf. ................................................................. 298

Health Gain, Cost Impacts, and Cost-Effectiveness of a Mass Media Campaign to Promote Smartphone Apps for Physical Activity: Modeling Study (e18014)
Anja Mizdrak, Kendra Telfer, Artur Direito, Linda Cobić, Tony Blakely, Christine Cleghorn, Nick Wilson. ................................................................. 313

Optimizing Smartphone-Delivered Cognitive Behavioral Therapy for Body Dysmorphic Disorder Using Passive Smartphone Data: Initial Insights From an Open Pilot Trial (e16350)
Hilary Weingarden, Aleksandar Matic, Roger Calleja, Jennifer Greenberg, Oliver Harrison, Sabine Wilhelm. ................................................................. 325

Evaluation of the Tobbstop Mobile App for Smoking Cessation: Cluster Randomized Controlled Clinical Trial (e15951)
Meritxell Pallejà-Millán, Cristina Rey-Reñones, Maria Barrera Uriarte, Esther Granado-Font, Josep Basora, Gemma Flores-Mateo, Jordi Duch. 3 3 6

Designing for the Co-Use of Consumer Health Technology in Self-Management of Adolescent Overweight and Obesity: Mixed Methods Qualitative Study (e18391)
Cynthia LeRouge, Hyeyoung Hah, Gloria Deckard, Haoqiang Jiang. ................................................................. 348
Feasibility and Acceptability of a Text Message-Based Intervention to Reduce Overuse of Alcohol in Emergency Department Patients: Controlled Proof-of-Concept Trial (e17557)
Elizabeth Burner, Mark Zhang, Sophie Terp, Kelsey Ford Bench, Joshua Lee, Chun Lam, Jesus Torres, Michael Menchine, Sanjay Arora.

Developing a Text Messaging Intervention to Reduce Deliberate Self-Harm in Chinese Adolescents: Qualitative Study (e16963)
Suqian Duan, Haoran Wang, Amanda Wilson, Jiexi Qiu, Guanmei Chen, Yuqiong He, Yuanyuan Wang, Jianjun Ou, Runsen Chen.

Use of a Machine Learning Program to Correctly Triage Incoming Text Messaging Replies From a Cardiovascular Text-Based Secondary Prevention Program: Feasibility Study (e19200)
Nicole Lowres, Andrew Duckworth, Julie Redfern, Aravinda Thiagalingam, Clara Chow.

Supplemental Text Message Support With the National Diabetes Prevention Program: Pragmatic Comparative Effectiveness Trial (e15478)
Natalie Ritchie, Silvia Gutiérrez-Raghunath, Michael Durfee, Henry Fischer.

Perceptions About Mindfulness and Text Messaging for Smoking Cessation in Vietnam: Results From a Qualitative Study (e17337)
Vuong Do, Claire Spears, Hoang Van Minh, Jidong Huang, Pamela Redmon, Nguyen Xuan Long, Michael Eriksen.

Describing the Process and Tools Adopted to Cocreate a Smartphone App for Obesity Prevention in Childhood: Mixed Method Study (e16165)
Paolo Giorgi Rossi, Francesca Ferrari, Sergio Amari, Andrea Bassi, Laura Bonvicini, Luca Dall’Aglio, Claudia Della Giustina, Alessandra Fabbri, Anna Ferrari, Elena Ferrari, Marta Fontana, Marco Foracchia, Teresa Gallelli, Giulia Ganugi, Barbara Ilari, Sara Lo Scocco, Gianluca Maestri, Veronica Moretti, Costantino Panza, Mirco Pinotti, Riccardo Prandini, Simone Storani, Maria Street, Marco Tamelli, Hayley Trowbridge, Francesco Venturelli, Alessandro Volta, Anna Davoli, Childhood Obesity Prevention Working Group.

User Experiences and Preferences Regarding an App for the Treatment of Urinary Incontinence in Adult Women: Qualitative Study (e17114)
Nienke Wessels, Lisa Hulshof, Anne Loohuis, Lisette van Gemert-Pijnen, Petra Jellema, Henk van der Worp, Marco Blanker.

A Mobile App Adopting an Identity Focus to Promote Physical Activity (MoveDaily): Iterative Design Study (e16720)
Floris Hooglugt, Geke Ludden.

A Scalable System for Passively Monitoring Oral Health Behaviors Using Electronic Toothbrushes in the Home Setting: Development and Feasibility Study (e17347)
Vivek Shetty, Douglas Morrison, Thomas Belin, Timothy Hnat, Santosh Kumar.

Limitations of Existing Dialysis Diet Apps in Promoting User Engagement and Patient Self-Management: Quantitative Content Analysis Study (e13808)
Jun-Hao Lim, Cordelia-Kheng-May Lim, Imluya Ibrahim, Jazlina Syahrul, Mohd Mohamed Zabil, Nor Zawaria, Zulfifri Daud.

Technologies for Opioid Use Disorder Management: Mobile App Search and Scoping Review (e15752)
Joseph Nuamah, Ranjana Mehta, Farzan Sasangohar.

Review and Selection of Online Resources for Carers of Frail Adults or Older People in Five European Countries: Mixed-Methods Study (e14618)
Roberta Papa, Areti Ethmiou, Giovanni Lamura, Flavia Piccinini, Giulia Onorati, Evridiki Papastavrou, Theologia Tsitsi, Giulia Casu, Licia Boccaletti, Alessandra Manattini, Rita Seneca, Carlos Vaz de Carvalho, Rita Durão, Francesco Barbabella, Frida Andréasson, Lennart Magnusson, Elizabeth Hanson.
Elderly Medication Adherence Intervention Using the My Interventional Drug-Eluting Stent Educational App: Multisite Randomized Feasibility Trial (e15900)
Andrew Boyd, Chioma Ndukwe, Anandu Dileep, Olivia Everin, Yingwei Yao, Betty Welland, Jerry Field, Matt Baumann, Jose Flores Jr, Adhir Shroff, Vicki Groo, Carolyn Dickens, Rami Doukky, Regeena Francis, Geraldine Peacock, Diana Willkie. 535

Influence of the SARS-CoV-2 Outbreak on the Uptake of a Popular Smoking Cessation App in UK Smokers: Interrupted Time Series Analysis (e19494)
Olga Perski, Aleksandra Herbe, Lion Shahab, Jamie Brown. 549

Developing Empirical Decision Points to Improve the Timing of Adaptive Digital Health Physical Activity Interventions in Youth: Survival Analysis (e17450)
Adrian Ortega, Christopher Cushing. 558

Enhanced Patient Education for Colonic Polyp and Adenoma Detection: Meta-Analysis of Randomized Controlled Trials (e17372)
Xu Tian, Ling-Li Xu, Xiao-Ling Liu, Wei-Qing Chen. 589

Influence of the Business Revenue, Recommendation, and Provider Models on Mobile Health App Adoption: Three-Country Experimental Vignette Study (e17272)
Francisco Lupiñez-Villanueva, Frans Folkvord, Mariek Vandenberg. 603

Feasibility of a Home-Based Tablet App for Dexterity Training in Multiple Sclerosis: Usability Study (e18204)
Judith van Beek, Erwin van Wegen, Marc Rietberg, Thomas Nyffeler, Stephan Bohihalter, Christian Kamm, Tobias Nef, Tim Vanbellingen. 616

The Ubiquitous Cognitive Assessment Tool for Smartwatches: Design, Implementation, and Evaluation Study (e17506)
Pegah Hafiz, Jakob Bardram. 625

The Mediating Role of Organizational Reputation and Trust in the Intention to Use Wearable Health Devices: Cross-Country Study (e16721)
Funmi Adebesin, Revingsome Mwalugha. 643

Continuous Measurement of Reconnaissance Marines in Training With Custom Smartphone App and Watch: Observational Cohort Study (e14116)
Leslie Saxon, Brooks D'Paual, Glenn Fox, Rebecca Ebert, Josiah Duhaime, Luciano Nocera, Luan Tran, Mona Sobhani. 658

A Nonproprietary Movement Analysis System (MoJoXlab) Based on Wearable Inertial Measurement Units Applicable to Healthy Participants and Those With Anterior Cruciate Ligament Reconstruction Across a Range of Complex Tasks: Validation Study (e17872)
Riasat Islam, Mohamed Bennasar, Kevin Nicholas, Kate Button, Simon Holland, Paul Mulholland, Blaine Price, Mohammad Al-Amri. 671

The Use of a Smartphone App and an Activity Tracker to Promote Physical Activity in the Management of Chronic Obstructive Pulmonary Disease: Randomized Controlled Feasibility Study (e16203)
Claire Bentley, Lauren Powell, Stephen Potter, Jack Parker, Gail Mountain, Yvonne Bartlett, Jochen Farwer, Cath O'Connor, Jennifer Burns, Rachel Cresswell, Heather Dunn, Mark Hawley. 687

Accuracy of Distance Recordings in Eight Positioning-Enabled Sport Watches: Instrument Validation Study (e17118)
Rahel Gilgen-Ammann, Theresa Schweizer, Thomas Wyss. 708

Testing Suicide Risk Prediction Algorithms Using Phone Measurements With Patients in Acute Mental Health Settings: Feasibility Study (e15901)
Alina Haines-Delmont, Gurjit Chahal, Ashley Bruen, Abbie Wall, Christina Khan, Ramesh Sadashiv, David Fearnley. 723
Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study (e19822)
Thomas Timmers, Loes Janssen, Joep Stohr, J Murk, M Berrevoets

Corrigenda and Addendas

Correction: Evaluating the Feasibility and Acceptability of a Mobile Health–Based Female Community Health Volunteer Program for Hypertension Control in Rural Nepal: Cross-Sectional Study (e19048)
Zhao Ni, Namratha Atluri, Ryan Shaw, Jingru Tan, Kinza Khan, Helena Merk, Yunfan Ge, Shrinkhala Shrestha, Abha Shrestha, Lavanya Vasudevan, Biraj Karmacharya, Lijing Yan

Correction: Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the "Needle in a Haystack" (e21600)
Florian Dittrich, Sascha Beck, Anna Harren, Felix Reinecke, Sebastian Serong, Jochen Jung, David Back, Milan Wolf, Stefan Landgraeb
Case-Initiated COVID-19 Contact Tracing Using Anonymous Notifications

Weibin Cheng1, MSc; Chun Hao2,3, PhD

1Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China
2Department of Medical Statistics and Sun Yat-sen Global Health Institute, School of Public Health & Institute of State Governance, Sun Yat-sen University, Guangzhou, China
3Health Information Research Center and Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou, China

Corresponding Author:
Weibin Cheng, MSc
Institute for Healthcare Artificial Intelligence Application
Guangdong Second Provincial General Hospital
466 Xingangzhong Road
Guangzhou, 510317
China
Phone: 86 20 89168139
Email: chwb817@gmail.com

Abstract

We discuss the concept of a participatory digital contact notification approach to assist tracing of contacts who are exposed to confirmed cases of coronavirus disease (COVID-19); the approach is simple and affordable for countries with limited access to health care resources and advanced technology. The proposed tool serves as a supplemental contract tracing approach to counteract the shortage of health care staff while providing privacy protection for both cases and contacts. This tool can be deployed on the internet or as a plugin for a smartphone app. Confirmed cases with COVID-19 can use this tool to provide contact information (either email addresses or mobile phone numbers) of close contacts. The system will then automatically send a message to the contacts informing them of their contact status, what this status means, the actions that should follow (eg, self-quarantine, respiratory hygiene/cough etiquette), and advice for receiving early care if they develop symptoms. The name of the sender of the notification message by email or mobile phone can be anonymous or not. The message received by the contact contains no disease information but contains a security code for the contact to log on the platform to retrieve the information.

(JMIR Mhealth Uhealth 2020;8(6):e20369) doi:10.2196/20369

KEYWORDS
COVID-19; surveillance; contact tracing; digital contact tracing; notification; anonymous; labor-saving; stigma; privacy protection

Introduction

Emerging evidence from the response to coronavirus disease (COVID-19) in China [1], Singapore [2,3], and South Korea [4] has indicated that efficient contact tracing reduces the delay between infection and isolation and accordingly prevents further transmission of the virus [5,6]. Contact tracing, which includes contact identification, listing, and follow-up, is a crucial aspect of epidemic control and is usually driven by health specialists. Contact tracing is a tedious task that requires enormous staff resources, and it is not possible to fully implement contact tracing in regions with widespread transmission. Alternatively, some parts of this task can be substituted and even augmented by technology.

Luca Ferretti and colleagues [1] modelled the potential effect of a digital contact tracing approach that involved training an artificial intelligence algorithm to analyze COVID-19 cases and GPS-based population colocalization information. However, the application of this powerful method may be hindered by advanced technical requirements and violation of privacy regulations. Yasaka and colleagues [7] developed a proof-of-concept smartphone app that allows users to create “checkpoints” for contact tracing and also modelled the effect of such an app under various adoption scenarios. This app respects user privacy by not collecting location information or other personal data. However, this approach relies on high levels of vigilance and willingness to participate among a majority of the population. Another concept of “privacy by design” COVID-19 contact tracing via Bluetooth is being rolled out in
Europe [8] which relies on Bluetooth data exchange between two mobile phones to detect whether two people have come into sufficient physical proximity to risk infection, and notifying those who have been in contact with an infected individual who stays anonymous. This type of app has the same constraint of requiring user cooperation to have any chance of success, and it may be more useful in developed countries where smartphones are widely used.

Health care staff resources have been scarce worldwide during the COVID-19 pandemic, especially in developing countries, where technological resources may also be inadequate. We discuss a concept of a contact notification tool to assist tracing of contacts who are exposed to confirmed cases of COVID-19; this tool is simple and affordable for countries with limited access to health care staff and advanced technology.

**Concept of the Tool**

The core functionality of our concept is to provide a usable, labor-saving tool for contact tracing by confirmed cases themselves (Table 1). This tool serves as a supplemental contract tracing approach to counteract the shortage of health care staff while providing privacy protection for both cases and contacts. This tool can be deployed on the internet or as a plugin for a smartphone app. Confirmed cases with COVID-19 can use this tool to provide contact information (either email addresses or mobile phone numbers) of close contacts; then, the system will automatically send a message to the contacts informing them of their contact status, what this status means, the actions that should follow (eg, self-quarantine, respiratory hygiene/cough etiquette), and advice for receiving early care if they develop symptoms. The name of the sender of the notification message by email or mobile phone can be anonymous or not. The message received by the contact contains no disease information but contains a security code for the contact to log on the platform to retrieve the information. This approach can prevent reading of the message by people other than the intended recipient. The personal identification data of both the confirmed case and their close contacts will not be recorded during the process. Information provided by confirmed cases will also be encrypted.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Descriptions and considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Products</strong></td>
<td>Smartphone app</td>
</tr>
<tr>
<td></td>
<td>Web-based platform</td>
</tr>
<tr>
<td><strong>Operating systems</strong></td>
<td>Windows and Apple</td>
</tr>
<tr>
<td></td>
<td>App compatible with the latest smartphone operating systems</td>
</tr>
<tr>
<td><strong>Target users</strong></td>
<td>Suspected and confirmed cases with COVID-19</td>
</tr>
<tr>
<td></td>
<td>People who receive messages from the platform</td>
</tr>
<tr>
<td></td>
<td>People who are concerned about COVID-19 infection</td>
</tr>
<tr>
<td></td>
<td>People who need guidance regarding the control and prevention of COVID-19</td>
</tr>
<tr>
<td></td>
<td>Policy makers and public health specialists</td>
</tr>
<tr>
<td><strong>Functionality</strong></td>
<td>Contact notification (eg, email, SMS text messages, autochatbot); can be anonymous or not</td>
</tr>
<tr>
<td></td>
<td>Shared exposure query database</td>
</tr>
<tr>
<td></td>
<td>COVID-19 information hub (optional but strongly recommended)</td>
</tr>
<tr>
<td><strong>Application scenarios</strong></td>
<td>Cases when in-person investigation cannot be implemented</td>
</tr>
<tr>
<td></td>
<td>Regions and countries where manual contact tracing is not in place</td>
</tr>
<tr>
<td></td>
<td>Supplement to manual contact tracing, especially in regions of widespread transmission where people with no or mild symptoms are self-quarantined at home</td>
</tr>
<tr>
<td><strong>Navigation and use of the tool</strong></td>
<td>Clear definition of close contacts of COVID-19</td>
</tr>
<tr>
<td></td>
<td>Guide for generation of close contacts</td>
</tr>
<tr>
<td></td>
<td>Official email account or telephone number</td>
</tr>
<tr>
<td></td>
<td>Template notification messages</td>
</tr>
<tr>
<td></td>
<td>Rights and responsibilities</td>
</tr>
<tr>
<td></td>
<td>Online consultation</td>
</tr>
<tr>
<td><strong>Other considerations</strong></td>
<td>Paramount protection of personal privacy</td>
</tr>
<tr>
<td></td>
<td>Strict adherence to general data protection regulations, careful oversight of data, and effective protections around the use of data</td>
</tr>
<tr>
<td></td>
<td>Guaranteed equity of access to testing and treatment</td>
</tr>
<tr>
<td></td>
<td>Careful design and monitoring to prevent malicious use</td>
</tr>
</tbody>
</table>

Table 1. Concept framework for the development of a tool for COVID-19 contact tracing.
Another key functionality is a shared exposure query database for people whose close contacts cannot be reached by messaging. This design is a crowdsourced database that allows patients to input their trajectory information, including public transportation used and times and locations of the patient’s movement prior to and after onset of symptoms. Trajectory information can either be provided by the confirmed patient or obtained during confirmation and uploaded by an epidemiologist. This trajectory information can be visualized on a map, which may be useful for the public to determine their possible exposure to COVID-19. The tool should be simple and easy to access so that it can be used in areas with limited resources. For example, people can visit the website and search by date and route for possible close contact with a patient with COVID-19. This tool can also serve as a hub to access instructions and information about COVID-19 and health services.

**Benefits of the Tool**

We envisage that this tool will be useful for people who are concerned about personal privacy and stigmatization related to COVID-19. Persons of Asian descent have faced stigma and discrimination in many places [9]. A person can also be subjected to stigma after they have been released from COVID-19 quarantine even though they are not considered to be at risk of spreading the virus to others. Several digital contact tracing approaches have been explored worldwide [10]. Despite the potential of digital contact tracing, its potential impact is limited because it may conflict with patient data privacy regulations. Our approach is different from other digital contact tracing methods in two ways. One is that it can be function as a website that does not need to be installed on the user’s smartphone. This will substantially lower the threshold for users to access the tool. The other is that no data processing, recording, or analysis occurs on the central server, which reduces concerns regarding personal privacy.

Meanwhile, this approach can be used as a supplement contact tracing tool in regions of widespread transmission where many undiagnosed mild cases are self-isolating at home, such as the United States and United Kingdom. This tool can help save health care resources, freeing staff to provide more urgent testing and clinical care for patients with COVID-19. Although mild cases who self-isolate at home may not associated with the spread of COVID-19, promptly informing contacts of possible risks to take proper precaution measures seems to be necessary and reasonable during a disease outbreak.

### Considerations for Successful Use

Several considerations should be taken into account to guarantee the successful use of this tool. First, the main purpose of this tool is to provide a supplementary approach for patients with COVID-19 to inform their friends, colleagues, and neighbors about possible contact while maintaining their privacy. Second, an accessible COVID-19 testing and care network is needed to guarantee equity of access to testing and treatment to meet the surging demand and relieve the anxiety of people who have received notifications and are anxious about possible infection. Third, privacy protection for both the sender and the recipient are equally important. The General Data Protection Regulation must be strictly followed, and careful oversight and effective protection of the use of data must be ensured. Last, we need to be alert to possible malicious use of this tool. People who do not have COVID-19 may use the system to send messages to others for evil purposes.

### Conclusion

The successful application of this tool relies heavily on public social responsibility and credibility, and it remains to be seen if the public would adopt such a tool and what mechanisms are required to prevent misuse. This is a simple tool that does not require complicated computer techniques despite strict user privacy protection design with respect to countries and regions. Additionally, this tool can help avoid coercive surveillance, facilitate the allocation of health resources, and prioritize clinical service for patients with COVID-19. Information obtained from the platform can also increase our understanding of the epidemiology of COVID-19.

### Conflicts of Interest

None declared.

### References


**Abbreviations**

*COVID-19*: coronavirus disease

©Weibin Cheng, Chun Hao. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 22.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Pantheoretical Framework to Optimize Adherence to Healthy Lifestyle Behaviors and Medication Adherence: The Use of Personalized Approaches to Overcome Barriers and Optimize Facilitators to Achieve Adherence

Azizi Seixas1*, PhD; Colleen Connors2, BA; Alicia Chung1, EdD; Tiffany Donley1*, DSC; Girardin Jean-Louis1*, PhD

1NYU Grossman School of Medicine, New York, NY, United States
2Georgetown University, Georgetown, NY, United States
*these authors contributed equally

Corresponding Author:
Azizi Seixas, PhD
NYU Grossman School of Medicine
180 Madison Avenue
New York, NY.
United States
Phone: 1 9728490862
Email: azizi.seixas@nyumc.org

Abstract

Patient nonadherence to healthy lifestyle behaviors and medical treatments (like medication adherence) accounts for a significant portion of chronic disease burden. Despite the plethora of behavioral interventions to overcome key modifiable/nonmodifiable barriers and enable facilitators to adherence, short- and long-term adherence to healthy lifestyle behaviors and medical treatments is still poor. To optimize adherence, we aimed to provide a novel mobile health solution steeped in precision and personalized population health and a pantheoretical approach that increases the likelihood of adherence. We have described the stages of a pantheoretical approach utilizing tailoring, clustering/profiling, personalizing, and optimizing interventions/strategies to obtain adherence and highlight the minimal engineering needed to build such a solution.

(JMIR Mhealth Uhealth 2020;8(6):e16429) doi:10.2196/16429

KEYWORDS
adherence; mHealth; management; chronic diseases; prevention; technology

Introduction

Background

Primary prevention vis-à-vis healthy lifestyle behaviors (healthy diet, nutrition, physical activity, sleep, and stress management) and management of chronic health conditions vis-à-vis pharmacological strategies significantly: (1) reduce the economic and medical burden of chronic diseases (eg, cardiometabolic disease and cancer) and corollary risk factors and (2) engender and improve positive health outcomes [1-8]. Poor adherence, the degree to which an individual does not follow pharmaceutical or lifestyle advice or regimen about primary prevention and management of chronic health conditions (such as lifestyle health behaviors and medications) [1] has significant economic and health consequences, resulting in greater health care expenditures, multiple morbidities, and deaths [9-11]. The burgeoning use of mobile technologies to deliver health, lifestyle and wellness interventions has shown initial signs of improving adherence to primary prevention and management of chronic health conditions [12-14]. However, the full potential of achieving optimized levels of adherence are thwarted by a wide range of sociodemographic, psychosocial, behavioral, and system-level barriers and the lack of personalized medicine and a precision population health approach—approaches that provide insight about the etiology of disease and health and promote customized, adaptive and just-in-time interventions based on biological/individual (eg, genes, biomarkers, circadian profile), lifestyle/behavioral (diet, physical activity, sleep, and stress management), and environmental/contextual (household, neighborhood, and cultural) factors. The purpose of this study was to explore (1) modifiable and nonmodifiable barriers and facilitators of adherence to primary prevention and management of chronic
health conditions, especially in mobile health (mHealth) solutions; (2) a precision and personalized population health framework that overcomes barriers and enables facilitators to adherence in primary prevention and management of chronic health, noncommunicable, and communicable health conditions; and (3) how to implement a precision and personalized population health approach in mHealth/digital health solutions.

Determinants of Health Adherence

Adherence to primary prevention and management of chronic health conditions in medicine and health is described as the degree to which an individual’s health and lifestyle behaviors are consistent with health recommendations [15,16]. Adherence can be influenced by (1) patient-related, (2) condition-related, (3) social and economic, (4) therapy-related, and (5) health system factors [17,18], which can be clustered into nonmodifiable (eg, race, gender, age) and modifiable (eg, health literacy, social support, and stress levels) factors. Distinguishing nonmodifiable and modifiable factors provides insights into how to modulate the effects of these barriers to improve adherence. For example, interventions based on and aimed at addressing nonmodifiable factors may not provide the best opportunities to improve and optimize adherence. Conversely, interventions based on and aimed at addressing modifiable factors may provide better opportunities to improve and optimize adherence to primary prevention and management of chronic health condition strategies.

Nonmodifiable Determinants of Adherence

Nonmodifiable factors are generally immutable and/or recalcitrant to change, which include sociodemographic (eg, age, race/ethnicity, sex) and psychosocial and behavioral (socioeconomic status [SES]) factors that individuals have little to no control over. Traditionally, adherence to primary prevention and management of chronic health condition solutions mistakenly target nonmodifiable factors to alter and increase adherence behaviors. Such an assumption would yield flawed assertions such as your race or sex determines your level of adherence to therapies and healthy behaviors. This should not be confused with acknowledging the likelihood that someone’s race or sex may be an important factor in their adherence behaviors. The distinguishing factor between the two approaches is that the former uses nonmodifiable factors to develop interventions that target subgroups, whereas the latter uses nonmodifiable factors to tailor interventions for subgroups of individuals that share common lived experiences (eg, neighborhood). For example, instead of targeting blacks, we can tailor interventions for low-income blacks living in midsized urban neighborhoods.

We acknowledge that there are rare times when nonmodifiable factors are the root cause of nonadherence behaviors, such as race-specific adverse medication side effects and suboptimal response to medications by certain racial/ethnic groups (eg, ACE inhibitors). Despite this, we argue that the fundamental premise of interventions aimed at increasing adherence behaviors is to modify factors that facilitate or impede adherence and to do this exclusively with nonmodifiable factors (eg, race, family history, and SES) is impossible. Instead, the process of tailoring interventions around nonmodifiable factors, whereby contents and activities of an intervention are geared toward cohorts of people classified by nonmodifiable factors (eg, race, family history, or SES) is a better strategy for obtaining adherence. Tailored solutions should use nonmodifiable factors to develop heterogeneous profiles of individuals to determine appropriate and congruent strategies to effect positive behavior change and maintain adherence to such behaviors.

Sex

Primary prevention and management of chronic health conditions (ie, medication and lifestyle behaviors adherence) vary by sex, where men and women display different levels of adherence across a wide variety of health behaviors and health conditions. Cross-sectional and longitudinal analyses of sex differences in medication adherence show consistently that women compared with men had less medication intensity, adherence to medications, and guidance and recommendations on drug use, especially among individuals with a chronic health condition [19,20]. Several studies indicate that men with HIV or diabetes were more adherent to physical activity recommendations and medication regimens compared with their female counterparts, partly due to greater levels of self-efficacy [21,22]. Conversely, other studies indicate that women typically reported higher health literacy levels, but lower on psychosocial determinants such as depression and social support, which are critical to medical adherence.

Age

Age is a significant predictor of adherence to primary prevention and management of chronic health conditions. Adherence levels across age groups are mixed, as some studies have shown that younger and older individuals display varying levels of adherence. In some studies, older adults reported greater levels of adherence, whereas in other studies, younger individuals had greater levels of adherence, depending on several factors such as the type of adherence behavior, health literacy, cultural beliefs, the personality of the individual, physical and cognitive impairment, self-perceptions of susceptibility, vulnerability and/or importance of health condition, and nature of chronic health conditions [23,24]. Studies also show that younger individuals compared with older individuals are more likely to be early adapters of certain treatments, while older individuals are more likely to demonstrate prolonged adherence. Generally, in web-based interventions, younger individuals have higher rates of intervention uptake compared with older individuals, and older individuals have higher levels of prolonging adherence compared with younger individuals [25].

Race/Ethnicity

Several studies have indicated that certain racial/ethnic groups have demonstrated varying levels of adherence to primary prevention and management behaviors of chronic health conditions. In a cross-sectional study of Medicare recipients living in Chicago, Gerber, Young, Ahsan, and Shou-Yih found that race influenced medication adherence, with elderly African American patients being less likely to follow physician instructions than their white counterparts, even after adjusting for potential confounding effects of depression, sociodemographic factors, health literacy, and social support [26]. Race may also affect adherence to healthy lifestyle
behaviors such as physical activity and dietary guidelines. Blacks and Mexican-Americans were less compliant with national physical activity/exercise recommendations of 150 min of moderate activity per week and reported higher levels of inactivity relative to white participants [27]. It should be noted that racial/ethnic differences in adherence are often confounded by several factors, such as access to health or medical resources, culture, income/SES, age, education level, language concordance, health literacy, and disability status [26-33].

**Socioeconomic Status**

Low SES (a combination of an individual’s education and income) and under-resourced communities typically affect adherence because they impede an individual’s ability to easily access quality and value-based health care. In addition, individuals in low-income and under-resourced neighborhoods are less likely to engage in healthy lifestyle behaviors such as healthy diet, physical activity/exercise, adequate sleep, and low stress [34,35]. Many low-income communities do not have easy access to gyms/fitness centers, doctors’ offices, healthy food grocers, parks and greenspaces, or pharmacies [36-41]. The lack of reliable transportation to commute to other neighborhoods that might have access to these health-promoting resources further compounds the dire nature of this situation [36,42]. In addition, low-income individuals with low health literacy may find it extremely difficult to adhere to health recommendations because they may not have the wherewithal and knowledge to advocate on their behalf to access quality and value-based health care [43].

**Modifiable Determinants of Adherence**

Compared with nonmodifiable determinants of adherence, modifiable determinants can be altered to increase adherence to primary prevention and management of chronic health conditions strategies. Some notable modifiable determinants of adherence include social support, motivation, emotional status, stress, health literacy, forgetfulness, health system, patient-provider communication, cost of health services, and health coverage and insurance.

**Social Support**

Social support is considered one of the most predictive modifiable factors in adherence to primary prevention and management of chronic health conditions. In a randomized control trial of 269 men and women aged 50 to 65 years, Oka, King, and Young found that support from friends, family, and exercise staff were the strongest predictors of adherence to maintaining physical activity and exercise regimen after 1 year [44]. Social support, via social media and network channels, is also integral in mHealth programs that attempt to increase adherence to primary prevention and management of chronic conditions such as physical activity, healthy diet, coping and stress reduction, and medication adherence [45,46].

**Motivation**

Motivation, which is the intrinsic or extrinsic driving force for initiating and maintaining goal-oriented behaviors, is another significant predictor of adherence to primary prevention and management of chronic health conditions. Motivation is key in initial uptake, adaptation, and maintenance of adherence behavior [47]. In chronic health conditions, motivation has proved effective in initial uptake and adaptation of health behaviors, but has shown mixed results in maintaining adherence behaviors, such as weight loss. Another challenge in optimizing motivation to increase adherence is to identify appropriate motivators for individuals across different contexts. For example, intended health benefits of adherence have proved to be insufficient in motivating individuals to adhere to healthy lifestyle prevention strategies in reducing the risk of obesity and cardiometabolic conditions [48]. Instead, other motivators such as incentives have been used successfully in mental health, home-based health monitoring, and exercise [49-51].

**Emotional Status and Stress**

Emotional status (depressive and anxiety symptoms) and stress can affect primary prevention and management of chronic health conditions, such as physical activity/exercise, sleep, and diet. Luyster, Hughes, and Gunstend conducted a cross-sectional study of 88 patients with heart failure and found that patients with symptoms of anxiety and depression are less likely to comply with their doctor-provided diet [52]. Perceived level of stress is another barrier, though more applicable to behavioral than medication adherence. In a cross-sectional study aimed at identifying barriers to exercise among women aged 40 years and older (745 African American, 660 Hispanic, 738 Native American/Native Alaskan, and 769 white), researchers found that individuals who reported feeling too tired for physical activity generally did so when they had a stressful day of work [53]. Stress has also been linked to poor adherence to healthy diets. Zellner and colleagues conducted an experiment to test the effect of stress on food choices among 34 female undergraduate students [54]. Participants were placed in a room with four different food choices (chips, peanuts, grapes, and M&Ms) and were asked to solve several problems with varying difficulty. The findings indicated that participants under more stress were more likely to eat M&Ms than healthier food options such as grapes. This concept known as emotional eating, due to high levels of distress, is associated with increased intake of high-calorie, low-nutrient foods, leading to weight gain and poor health outcomes [55-57]. Generally, these associations are not putative and are often mediated or confounded by nonmodifiable factors such as sex/gender. In a study that investigated self-management behavior among individuals with diabetes, men with lower levels of depression and anxiety displayed greater levels of self-management and adjustment to disease-related challenges relative to their female counterparts [58].

**Health Literacy**

Over 90 million Americans report inadequate literacy about healthy behaviors and lifestyle. Inadequate health literacy may compromise adequate comprehension of primary prevention and management of chronic health conditions. Muir et al [59] found that individuals with higher levels of health literacy are more likely to adhere to health recommendations than those with lower levels of health literacy. Gazmararian et al [60] also found that individuals with inadequate health literacy had greater odds of low refill medication adherence than those with adequate health literacy. Health literacy may also be affected by nonmodifiable and modifiable determinants of adherence.
behaviors, such as age, education level, cognitive impairment, dosing frequency, and patient-related health concerns [23,61-64].

**Cognitive Factors (Memory and Information Processing)**

Memory impairment and difficulty in processing health information and instructions are associated with poor adherence to primary prevention and management of chronic health conditions. Patients forget 40% to 80% of their health information and instructions given to them by their health care providers, thus compromising their ability to prevent or manage their risk for chronic health conditions [65]. In one study, impairment to prospective memory accounted for 24% of nonadherence to primary prevention and management of chronic health conditions [66]. Prospective memory is defined as an individual’s ability to remember to do something at a later time and is considered one of the strongest predictors of nonadherence [67-69]. To address memory-related nonadherence, providing cues and reminder alerts have helped individuals successfully achieve adherence to primary prevention and management of chronic health conditions [69,70].

**Health Systems**

Value-based health care (the ability to access quality and affordable health care) and navigating the complex and inaccessible health system are two system-level barriers that impede adherence to primary prevention and management of chronic health conditions [71]. Expensive health care systems, navigating complex health insurance and payer systems, limited operation hours of health facilities, and difficulty in navigating the complex health system are significant health system barriers to adherence [71]. Poor and inadequate patient-provider communication has proved to be another significant barrier to primary prevention and management of chronic health conditions. In a study of 5929 patients across 13 different hospital systems, patients with inadequate health literacy reported poorer patient-centered communication [72]. Poor patient-centered communication may be a derivative of the current payment structure in health care, which limits and disincentivizes the length of time patients have with providers. Cost and financial stress are additional health system barriers to adherence [73-76]. For example, in a systematic literature review, higher out-of-pocket medical costs were associated with a decrease in medication adherence [77]. Inefficiencies in the delivery of health care services, especially interdepartmental care coordination, are another significant health system barrier to adherence. The World Health Organization in a systematic review reports that poor provider communication about follow-up plans (eg, discharge and continuation plan), side effects of treatment, treatment journey and trajectory with patients and other providers (clinicians and pharmacists), poor information technology infrastructure, and poor multidisciplinary treatment team infrastructure make it difficult for patients to adhere to recommended treatment and medical advice [9].

**Overcoming Nonmodifiable and Modifiable Barriers to Health-Related Adherence**

To overcome nonmodifiable and modifiable barriers of health-related adherence behaviors requires solutions that are both nomothetic/one-size-fits-all and idiographic/personalized. Such solutions must embrace a pantheoretical approach, one that incorporates nomothetic and idiographic approaches to engender precise, personalized, and optimized (contextualized) solutions to increase adherence to primary prevention and management of chronic health conditions. Nomothetic approaches are characterized as group-based or tailored interventions that apply to all and are generally based on nonmodifiable factors. However, idiographic approaches are characterized as precise and personalized interventions generally based on modifiable factors (Figure 1).

The pantheoretical approach consists of four major processes: tailoring, clustering, personalization, and optimization (Figure 1). Traditionally, tailoring is considered to be the act of customizing treatments for certain groups, based on age, race/ethnicity, and location (eg, urban/rural). However, in the pantheoretical approach, tailoring entails identifying nonmodifiable determinants of adherence behaviors and creating educational content based on profiles of nonmodifiable factors (eg, a health education program targeting black women or Hispanic men). Tailored approaches have proved effective in obtaining adherence across a wide variety of health outcomes [78,79].

Clustering, the second phase in the pantheoretical approach, consists of identifying modifiable determinants of adherence (generally behavioral or psychological), creating behavioral profiles from these nonmodifiable factors, identifying which factor(s) should be modified to increase adherence for an individual, and providing information/messages to address specific modifiable factor(s) responsible for nonadherence. The factors that are intervened upon are considered active ingredients in behavior change.

Personalization, which is the third phase, consists of developing a meta-cognitive/mind map of barriers and facilitators of adherence (Figure 2) [80,81]. This meta-cognitive/mind map can be developed from qualitative and/or quantitative data, specifically from likelihood estimates from either trials, prospective studies, or meta-analyses. The first step in developing the meta-cognitive/mind map is to draw the conceptual model with all factors and how they are related to each other via arrows. The second step entails parameterizing each relationship. Typically, parameterization involves affixing weights based on the relative importance of each factor and creating quantitative equations that represent each relationship among factors in the meta-cognitive/mind map. If data are not available, the meta-cognitive/mind map can be derived from iterative consensus building (generally through DELPHI-like focus groups) and affixing numerical weights (representing relative importance of factors) and defining relationships among factors through hypothesized mathematical formulas based on extant research. It is highly recommended that simulation modeling software tools are used to test, validate, and calibrate these models. Meta-cognitive/mind map models can facilitate the initial development of an algorithm that can be utilized in interventions.

Optimization, the fourth phase, is the amalgamation of a tailored, clustered, and personalized intervention. A fully optimized
intervention is adaptive, provides information, and messages just-in-time with an appropriate dosage at the right time in the right context [82]. To achieve complete optimization, several iterations, experiments, and fine-tuning are needed.

**Figure 1.** Workflow and framework to incorporate nonmodifiable and modifiable factors to improve and optimize health-related adherence behavior.

### Nonmodifiable Determinants of Adherence
- Age
- Sex/Gender
- Race/ethnicity
- Socioeconomic status
- Place of residence
- Culture
- Physical ability status
- Family history/Genetics

### Modifiable Determinants of Adherence
- Social support
- Emotional Status and Stress
- Health Literacy
- Patient Cognition (memory and information processing)
- Health system
- Health Insurance

### Tailoring (Group, nomothetic and segmentation profiling)
- Assess and identify non-modifiable determinants to adherence.
- These factors will be used to profile individuals and will not be primary focus for interventions to increase/optimize adherence.
- Non-modifiable factors will help to tailor interventions.

### Clustering (Idiographic Profiles)
- Assess and identify modifiable (typically behavioral and psychological) determinants to adherence.
- These factors will be used to create behavioral profiles to determine which factors to intervene on to increase/optimize adherence.
- Modifiable factors are active ingredients in intervention

### Personalize (Group + Idiographic profiles)
- For studies without large dataset to machine learn profiles construct meta-cognitive map for adherence behavior (non-modifiable and modifiable) based on meta-analyses and extant studies (Top-down)
- For studies with data, use machine learning to develop cluster profiles of non-modifiable and modifiable factors and then determine which profile a patient fits based on their baseline data (Bottom-up)

### Optimization
- Optimization of adherence = Tailoring + Clustering + Personalization + Context
- Iterate, experiment, and learn ideal context to optimize adherence behavior
- Develop an adaptive, just-in-time, just-enough dosage, real-time intervention with the use of cloud computing
Developing a Fully Personalized and Optimized Intervention

Personalization is defined as the incorporation and analysis of personal, behavioral, and real-time data to create highly contextual, adaptive, and automated messaging and curriculum (videos, text, and the literature) relevant to the participant’s emerging needs that can maximize behavior change. It is widely used in market research and is successful in enhancing engagement, experience, acceptance, and adherence to product recommendations. We argue that personalization is an effective strategy for changing and maintaining health behavior. However, to develop and achieve evidence-based personalization and optimization strategies, four critical steps must be taken.

**Step 1**
Develop a detailed idiographic profile represented as a meta-cognitive/mind map (Figure 2, idiographic profile and care journey of chronic disease) using an individual’s behavioral (behaviors that impact health and adherence to healthy lifestyle and wellbeing, such as sleep, physical activity, diet, and stress), demographic (eg, age, sex, and race/ethnicity), psychosocial (neighborhood and household context), clinical (any medical diagnosis or risk of disease), and cultural (attitudes, beliefs, and cognitions about health and wellness) data. Idiographic profiles consist of modifiable and nonmodifiable determinants of the 6 As of the chronic disease care journey: awareness, avoidance, access, assessment, acceptance, and adherence. The goal of a personalized approach is to target modifiable determinants (such as attitudes, behaviors, beliefs, and cognitions) to increase the likelihood of awareness, avoidance, access, assessment, acceptance, and adherence behaviors.

**Step 2**
Develop a generic health care-continuum journey map to achieve adherence for a specific health condition via focus groups from multiple stakeholders. The care journey map should identify critical bottlenecks and decision points (moments where participants are likely to relapse or adhere) in achieving treatment adherence as well as potential modifiable factors (intentions, behaviors, and motivations) for each bottleneck and decision point. As key modifiable factors for each bottleneck and decision point are idiosyncratic, developers should capture as much data on a wide range of potential modifiable factors before intervention deployment.

**Step 3**
Develop and train a prediction algorithm that can identify critical decision points (vulnerable and opportune times to adhere) based on the individual’s idiographic profile and care continuum journey map. This is best achieved by 1 of 2 strategies: (1) an *a priori* approach (profiling the individual via baseline data collection and determining if their idiographic profile fits a previously validated profile) and (2) a *a posteriori* approach (developing an idiographic profile via baseline data or a phase-in stage where the individual is observed without being exposed to any intervention). To validate decision points, it is important to capture the input of patients in real time via responses to queries from ecological momentary assessments to maximize the timing, frequency, and duration of the intervention [83,84].

**Step 4**
After validation of decision points and bottle necks, the care journey must undergo further refinement. Beyond the initial stages of intervention exposure, the idiographic profile of each individual must be constantly updated to reflect any changes in their profile. To do so, we will capture in real time, participants’ responses to ecological momentary assessments to maximize timing, frequency, and duration of intervention exposure over a prolonged period. These data will be consistently ingested, stored, and analyzed through cloud computing. Insights obtained from cloud computing algorithms will trigger the delivery of adaptive and personalized behavioral interventions to sustain healthy behaviors. In summary, the proposed personalized models will amalgamate patient-level factors with real-time changes to create dynamic idiographic profiles, enabling the delivery of just-in-time messages that are responsive to real-time context.

Implementing and Testing the Pantheoretical Approach to Behavior Change Through Mobile Health Technology

To successfully achieve all the components of the pantheoretical approach (tailoring, clustering, personalization, and optimization), a method that is accessible, portable, and nimble (easily modifiable and adaptable), such as mHealth technology,
is needed. Embedding the pantheoretical framework in mHealth solutions can mitigate barriers and help optimize facilitators of adherence to primary prevention and management of chronic health conditions. A mHealth solution is an ideal platform and medium to achieve and sustain health behavior change through personalized and optimized strategies.

How Can Mobile Health Help?

Despite the proliferation of mHealth solutions and the high uptake and short-term use of such solutions, prolonged use is low and compromises long-term adherence. To engender long-term use of mHealth, it is important to address modifiable and nonmodifiable barriers. Although mHealth solutions cannot fully address the inhibitory effects of nonmodifiable barriers to adherence, they offer real and novel strategies to overcome and circumvent modifiable barriers. Some of these barriers are poor engagement, lack of motivation, limited support to achieve healthy behavior change across different contexts, inadequate health literacy, and limited access to evidence-based strategies to support healthy behavior change in real time.

One major patient-level modifiable barrier is lack of motivation, as patients may find it difficult to complete recommended tasks to achieve healthy behaviors and lifestyle. mHealth tools can modify motivation by providing timely motivators (eg, rewards, conditioned stimuli, and reinforcers) and information that may activate intrinsic beliefs and attitudes and thus induce positive behavior change. One system-level modifiable barrier is access to evidence-based strategies to support healthy behavior changes in real time. Unlike a health care provider, mHealth solutions can provide ubiquitous support (eg, out of office) to patients via mobile devices. Mobile solutions can be easily integrated into a patient’s daily routine to increase adherence to medical treatment and advice and primary prevention and management of chronic health conditions.

First, mHealth solutions can make adherence easier and more user-friendly for patients by reducing the cognitive demand and fatigue required for optimal adherence. Specifically, mHealth solutions can optimize prospective memory and information processing, improve health literacy, reduce recall bias on health behaviors and lifestyle practices, provide social support through networks of patients with similar conditions, and facilitate effective and timely patient-provider communications. Second, mHealth solutions can also address psychosocial and health system barriers to adherence by reducing the cost of health care and increasing access to specialized and expensive health care providers. Third, mHealth solutions also allow patients to monitor their health on their own time, and access and communicate with health care providers at convenient times.

Health care providers also benefit from the use of mHealth solutions. Since providers are heavily reliant on subjective and sometimes unreliable patient reporting, an mHealth solution would allow providers to continuously monitor patient performance and adherence, reducing the likelihood of recall bias, and increasing the possibility of more personalized and adaptive interventions. mHealth solutions allow providers to be more accessible to their patients with less effort to provide real-time support through several engineering technologies, such as push-notification reminders and alerts to comply with recommended treatments, automated bots with artificial intelligence knowledge banks that can provide dynamic information to patients based on their unique conditions, and effective strategies and tips to optimize medication and healthy behavior/lifestyle adherence through the internet of things things/devices that independently or jointly with other things/devices collect, store, and process data via digital or sensor-based technology, for example, wearable sensors.

The development of such a health service delivery system requires novel engineering and a paradigmatic shift in practicing medicine from a warehouse, a one-size-fits-all approach to a personalized and optimized system. An mHealth solution rooted in a pantheoretical framework (Figure 1) can revolutionize the delivery of health services from a one-size-fits-all to a personalized and optimized approach. A personalized and optimized approach can optimize adherence to prevention strategies and management of chronic health conditions, which will go at the patient’s pace, thus optimizing the likelihood of adhering to recommended medical treatment and advice, making health behavior goals more achievable.

Who Will Benefit From a Personalized and Optimized Mobile Health Solution?

Despite the potential benefits and successes of a personalized and optimized mHealth solution for adherence, there are some people for whom this strategy may not work. Therefore, it is important to identify those patients who are ideal candidates. There are two strategies that can be employed to determine a patient’s candidacy. The first approach utilizes an a priori strategy to generate a behavioral profile that includes the patient’s likelihood of being engaged, ready to participate, and adhere to an mHealth solution. The second strategy utilizes a trial-and-error run-in phase. This strategy exposes patients to the mHealth solution and monitors their level of engagement, participation, and adherence, and if after a certain period, the patient is not responsive, then it is likely that they may not be a good candidate. However, if it is critical that the patient participates in this mHealth solution, it is highly recommended to enroll the patient in a pretreatment preparatory training, which will help them be more responsive to an mHealth solution. For example, individuals above 65 years (older adults), with a chronic health condition, low-income background, racial/ethnic minority background, or inadequate health literacy are traditionally not ideal candidates for an mHealth solution [85]. Unfortunately, these are the people who need a personalized and optimized mHealth solution most. Therefore, pretreatment preparatory training would entail identifying an individual’s unique barriers to engagement with mHealth solutions and creating a graduated and sequential curriculum that involves education, simulated, and real-world use of mHealth solutions.

Limitations to Mobile Health Solutions: Technical, Psychosocial, and Financial Barriers

Technical, psychosocial, and financial factors that affect the adoption and prolonged use of mHealth are often overlooked yet critical barriers. Users’ reluctance to embrace mHealth solutions may include fear of technology, inability to purchase, download, and navigate mHealth and digital apps and solutions, and overly stimulating and busy interfaces making following
instructions and user experience and engagement more challenging. These barriers might explain why in general one-third of individuals stop using mobile and digital solutions within 6 months of download [86].

Cost, access, privacy concerns, user satisfaction, technology literacy, and proficiency are barriers to mHealth use. Despite the fact that 1 in 5 Americans use an mHealth solution, prolonged use depends on user satisfaction, learnability, efficiency, errors, and memorability [83,84]. Data privacy and security may be other areas of concern, particularly for shared personal health information [87]. Barriers to mHealth adoption may also be due to cost, given the moderate to high-income user base [84]. The cost of a mobile device and monthly telecommunication service charge (the traditional model in the United States) serve as two major barriers. Finally, many mHealth solutions require the purchase of expensive wearable or fitness tracking devices for real-time tracking and sophisticated data visualization of user performance. Individuals are often priced out of mHealth solutions, thus giving rise to health technology inequities, where only upper- and middle-class individuals will likely benefit. Despite these limitations, the proliferation of cheaper and more affordable mobile devices as well as their use in adjunctive services provided by health care facilities and payers in comprehensive prevention health programs serve as silver linings.

Conclusions

There are several factors that affect an individual’s adherence to health recommendations. As described above, these factors can be categorized into two categories: nonmodifiable and modifiable factors. Nonmodifiable factors are difficult to change and influence; thus, they may not provide the best opportunities to affect behavior change that leads to long-term adherence to primary prevention and management of chronic health conditions. Therefore, we argue that addressing modifiable determinants such as social support, health literacy, user motivation, emotional status, cognition (memory and information processing), and healthcare systems may provide better opportunities to affect behavior change and long-term adherence to health behaviors. We further argue that a mHealth solution may be a viable approach to address modifiable barriers and optimize adherence, while taking into consideration nonmodifiable factors, which serve to tailor, cluster/profile, personalize, and optimize interventions/strategies to obtain adherence, the pantheoretical approach.

Although mHealth solutions can be ideal for successful achievement and maintenance of adherence behaviors, they can also exacerbate barriers and thus compromise adherence. For example, low-income individuals who cannot afford mHealth solutions may be prohibited from accessing mHealth solutions, thus increasing the likelihood of nonadherence and unhealthy behaviors. In addition, the low rates of prolonged use of mHealth solutions is another critical barrier that must be addressed if it is to be used for long-term adherence. We argue that to maximize the full potential of mHealth solutions to obtain and maintain adherence, developers need to create more engaging, personalized, tailored, and multidimensional solutions (those that take into consideration the role of nonmodifiable and modifiable determinants on adherence) to achieve long-term adherence.

Acknowledgments

This research was supported by funding from: K01HL135452, R01MD007716, R01HL142066, T32HL129953, and K07AG052685.

Conflicts of Interest

None declared.

References


44. Oka RK, King AC, Young DR. Sources of social support as predictors of exercise adherence in women and men ages 50 to 65 years. Womens Health 2019 [Medline: 39373378]


Abbreviations

**mHealth**: mobile health

**SES**: socioeconomic status

©Azizi Seixas, Colleen Connors, Alicia Chung, Tiffany Donley, Girardin Jean-Louis. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 24.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Medical Device Apps: An Introduction to Regulatory Affairs for Developers

Abstract

The Poly Implant Prothèse (PIP) scandal in France prompted a revision of the regulations regarding the marketing of medical devices. The new Medical Device Regulation (MDR [EU]) 2017/745 was developed and entered into force on May 25, 2017. After a transition period of 3 years, the regulations must be implemented in all EU and European Economic Area member states. The implementation of this regulation bears many changes for medical device development and marketing, including medical device software and mobile apps. Medical device development and marketing is a complex process by which manufacturers must keep many regulatory requirements and obligations in mind. The objective of this paper is to provide an introduction and overview of regulatory affairs for manufacturers that are new to the field of medical device software and apps with a specific focus on the new MDR, accompanying harmonized standards, and guidance documents from the European Commission. This work provides a concise overview of the qualification and classification of medical device software and apps, conformity assessment routes, technical documentation, clinical evaluation, the involvement of notified bodies, and the unique device identifier. Compared to the previous Medical Device Directive (MDD) 93/42/EEC, the MDR provides greater detail about the requirements for software qualification and classification. In particular, rule 11 sets specific rules for the classification of medical device software and will be described in this paper. In comparison to the previous MDD, the MDR is more stringent, especially regarding the classification of health apps and software. The implementation of the MDR in May 2020 and its interpretation by the authorities will demonstrate how app and software manufacturers as well as patients will be affected by the regulation.

Introduction

Due to safety issues in the field of medical devices, and especially after the Poly Implant Prothèse (PIP) scandal in France, the Medical Device Directive (MDD) 93/42/EEC [1] was revised and replaced with the new Medical Device Regulation (MDR [EU]) 2017/745 [2,3]. The MDR entered into force on May 25, 2017 and must be implemented within the European Union (EU) and European Economic Area (EEA) states after 3 years, by May 26, 2020 [2]. Since the MDR is a regulation, it is immediately enforceable as law in all member states after its implementation date. This contrasts with the previous MDD, which was a directive that member states transpose into national law within a set timeframe [4].

Unlike the US Food & Drug Administration, which regulates foods, medicines, and medical devices, the European Medicines Agency (EMA) regulates only drugs. There is no regulatory body like the EMA for the review and approval of medical devices. Manufacturers themselves declare conformity of their devices with the European legislations and regulations and affix a CE (Communauté européenne) mark (Article 10 and 20 MDR). Products that bear a CE mark can then be marketed within the EU/EEA (Articles 2 and 10 MDR). Affixing the CE mark to a product is only legal after a conformity assessment has been
performed (Article 20 MDR). Depending on the class of the device, a notified body must be involved in this process (Introduction [60] and Articles 52-53 MDR). For certain highly critical or novel products, an additional examination by so called “expert panels” is mandatory (Introduction [56] and Article 106 MDR).

Independent of the risk class, technical documentation (TD) must be compiled to allow an assessment of whether the general safety and performance requirements set by the MDR are met (Annex II MDR). With the exception of class I devices, the notified bodies then inspect the manufacturer’s Quality Management System (QMS) and technical documentation and subsequently issue the required Annex certificates (Annex XII MDR). These are prerequisites for declaring conformity and affixing the CE mark (Article 10[6] MDR). For guidance on how to perform the steps required for CE marking, including risk management, technical documentation, and QMS, and to prove regulatory compliance, manufacturers are advised to work according to harmonized standards and common specifications (Articles 8-9 and Annex II [4c] MDR). When the MDR came into force in 2017, there was no associated harmonized standard or common specification. According to the European Commission, these will be implemented soon [5].

When developers of software or mobile apps claim that their product has a medical purpose, it becomes a medical device and must bear a CE mark (Article 2[1] MDR). This paper describes the process of placing mobile apps and software on the market as medical devices (Figure 1) and serves as an introduction to regulatory affairs for app and software developers.

Figure 1. Important stages of medical device development.

Elements of the Medical Device Regulation

The primary elements of the new Medical Device Regulation (MDR [EU]) 2017/745 and the accompanying harmonized standards and guidance documents provided by the European Commission are described below.

Qualification of Mobile Apps and Software: What Constitutes a Medical Device?

Medical Device Software

Mobile apps and software that are independent of any device and are not intended to be used as an accessory to a medical device are referred to as Medical Device Software (MDSW) or standalone software [6] and must be qualified and classified in their own right (Annex VIII [3.3] MDR).

Intended Purpose

The first, essential question an app or software developer must answer is whether the product is a medical device or not. Software qualifies as a medical device if the developer’s stated purpose of the software meets the definition of a medical device in Article 2[1] of the MDR (Textbox 1).

Textbox 1. Medical device definition [Article 2(1) MDR].

“Medical device” means any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or more of the following specific medical purposes:

- diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease,
- diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,
- investigation, replacement or modification of the anatomy or of a physiological or pathological process or state,
- providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations, and which does not achieve its principal intended action by pharmacological, immunological or metabolic means, in or on the human body, but which may be assisted in its function by such means. The following products shall also be deemed to be medical devices:
  - devices for the control or support of conception;
  - products specifically intended for the cleaning, disinfection or sterilisation of devices as referred to in Article 1(4) and of those referred to in the first paragraph of this point.
The decision of whether a software product qualifies as a medical device is made by the developer or, using the terminology of the MDR, the manufacturer (Article 2[30] MDR). If the manufacturer states that the system can be used for a medical purpose, it must be CE marked (Articles 10 and 20 MDR). Whether a product qualifies as a medical device is determined by the intended use, as stated by the manufacturer and the mechanism of action of the product, not the design or user [6]. Furthermore, the description of the intended purpose must include a statement of benefit for the patient, otherwise it cannot be marketed as a medical device (Articles 61-62, Annexes XIV and XV MDR).

**Key Characteristics for Qualification as a Medical Device**

If a mobile app or software performs an action on data beyond storage, archiving, communication, or simple search; the performed action is for medical purposes; and the performed action is for the benefit of an individual patient, it most likely qualifies as a medical device and is subject to the MDR [6]. For further guidance on qualification of standalone software, see “MEDDEV 2.1/6 Guidelines on the qualification and classification of standalone software” [7].

### Classification of Standalone Software Including Apps

The MDR defines four risk classes: I, IIA, IIB, and III (Table 1). For classification of software, rule 11 has been included in the regulation (Textbox 2) (Annex VIII MDR).

<table>
<thead>
<tr>
<th>Class</th>
<th>Documentation Requirement</th>
<th>Notified body involved?</th>
<th>QMSa</th>
<th>Certificates</th>
<th>Clinical investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (low risk)</td>
<td>Manufacturer must compile the technical documentation and self-declare conformity</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Not mandatory. May be required depending on the outcome of the clinical evaluation</td>
</tr>
<tr>
<td>IIA (low-medium risk)</td>
<td>Manufacturer must draw up the technical documentation and apply to a European Notified Body</td>
<td>Yes</td>
<td>Yes, certified</td>
<td>Yes (Annex IX certificate, QMS certificate)</td>
<td>Not mandatory. May be required depending on the outcome of the clinical evaluation</td>
</tr>
<tr>
<td>IIB (medium-high risk)</td>
<td>Manufacturer must draw up the technical documentation and apply to a European Notified Body</td>
<td>Yes</td>
<td>Yes, certified</td>
<td>Yes (Annex IX certificate, QMS certificate)</td>
<td>Not mandatory. May be required depending on the outcome of the clinical evaluation</td>
</tr>
<tr>
<td>III (high risk)</td>
<td>Manufacturer must draw up the technical documentation and apply to a European Notified Body</td>
<td>Yes, expert panel</td>
<td>Yes, certified</td>
<td>Yes (Annex IX certificate, QMS certificate)</td>
<td>Mandatory</td>
</tr>
</tbody>
</table>

aQMS: Quality Management System.

---

**Textbox 2. Rule 11, Annex VIII MDR.**

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIA, except if such decisions have an impact that may cause:

- Death or an irreversible deterioration of a person's state of health, in which case it is in class III; or
- Serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIB.

Software intended to monitor physiological processes is classified as class IIA, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIB. All other software are classified as class I.

This rule implies that many apps might have to be classified as class IIA, IIB, or III in the future, while under the former Medical Device Directive (MDD) 93/42/EEC most standalone software including apps were classified as class I or not designated as medical devices at all [1]. Interpreting rule 11 on its own indicates that, for example, software used to calculate dosages of drugs with high toxicity, suggesting a diagnosis, or aiding with therapy or radiation planning will fall within class III, since a mistake might cause death. If it is highly unlikely that death could be caused by an error, it could fall within class IIB, which is defined as a device for which “...a mistake can cause serious deterioration of a person’s state of health...” (Annex VIII MDR). Medical device software may only fall under class IIA if a mistake cannot be anticipated to cause serious deterioration of a person’s state of health. Medical device software may fall under class I only if it is not intended to provide information used to make decisions for diagnostic or therapeutic purposes. For class I devices, no notified body is involved in the declaration of conformity.

These classification criteria are very stringent and prompted formation of the Medical Device Coordination Group (MDCG), an official working group serving to advise the European Commission regarding medical devices [8]. The group recently released a guidance document that further elaborates rule 11 [6]. The guidance provides examples on how to classify...
Conformity Assessment Routes

Fulfilling the “General safety and performance requirements” described in Annex I of the MDR is the most crucial step on the long road to marketing a medical device. To prove compliance with these requirements, manufacturers must follow one of the conformity assessment procedures described in the MDR appendices. Typically, manufacturers apply harmonized standards, and in the future also common specifications, to prove compliance with these requirements (Articles 8-9 and Annex II [4C] MDR). Examples of general safety and performance requirements include risk management, software lifecycle processes, software verification and validation, and usability (Annex I MDR). A complete list of requirements can be found in Annex I of the MDR.

Conformity assessment is a process demonstrating whether the “general safety and performance requirements” of the MDR (Annex I MDR) have been fulfilled. Once the conformity of the medical device with the requirements has been proven, the manufacturer may declare the conformity, CE mark and market the product within the EU/EEA (Articles 19-20 MDR). Depending on the risk class, the MDR describes three different paths of conformity assessment in accordance with Annexes IX, X and XI (Article 56 MDR). Besides the four main risk classes (I, IIa, IIb, III), the MDR defines three sub-classes for risk class I, which are devices with measuring function (I_m), sterile devices (I_s) and reusable surgical instruments (I_r). For classes I_m, I_s, I_r, IIa, IIb and III, a notified body must be involved in the process of conformity assessment. This is not required for all remaining devices falling within class I (Articles 52-53 MDR). For risk class III, an additional expert panel will scrutinize the clinical evaluation and assess whether the clinical data is sufficient to provide confidence in the safety and performance of the device (Annex IX MDR). For medical device software, the development process cannot be ignored since it is difficult to find errors by simply testing the finished product, which would be the case in the Annex XI (Part B) conformity assessment. In contrast, the procedure described in Annex IX includes an assessment of the QMS and the technical documentation by a notified body. The product of a successful assessment is an Annex IX certificate and an EU QMS certificate for the manufacturer, who can then declare conformity of the medical device with the requirements set by the MDR (Articles 19 and 56 MDR).

Technical Documentation

According to the “general obligations of manufacturers” (Article 10[4] MDR), technical documentation must be compiled and kept up to date to enable assessment of compliance with the safety and performance requirements set by the MDR. Annex II of the MDR lists in detail what is required in the documentation, including documentation of a fully implemented risk management system, benefit-risk analysis, clinical evaluation report, software life cycle file, usability file, and many other requirements.

Clinical Evaluation

The supporting documentation for the CE declaration must include a clinical evaluation. This is an evaluation of side effects and the acceptability of the benefit-risk ratio based on clinical data (Article 61 MDR). Conducting a proper clinical evaluation will demonstrate (1) which clinical data are necessary; (2) which clinical data can be adequately supplemented by methods other than clinical investigations, such as published literature, prior clinical investigations, clinical experience, or by using suitable clinical data from equivalent devices; and (3) which clinical data remain to be delivered by clinical investigations (Article 61 MDR). Clinical investigations within the MDR are what many people would refer to as “clinical trials” and are defined as “systematic investigation involving one or more human subjects, undertaken to assess the safety or performance of a device” (Article 2[45] MDR). It is one of the methods to obtain clinical data supporting treatment efficacy and to confirm clinical benefit (Article 62 MDR). If sufficient clinical data to perform a clinical evaluation can be retrieved from the literature or other sources, the manufacturer can proceed without a clinical investigation. The clinical evaluation must be updated frequently with data from post-market surveillance. New data, as well as considerations for new or changed intended purposes, require an updated clinical evaluation and may indicate the necessity for additional clinical investigations (Article 61 MDR).

The clinical evaluation must be planned, conducted, and documented. The clinical evaluation, its results, and the clinical evidence derived from it must be documented in a clinical evaluation report and included as part of the technical documentation (Annexes II and XIV MDR[9]). Furthermore, a clinical evaluation plan must be elaborated and documented in the technical documentation (Annexes II and XIV MDR).

Regulatory Requirements for the Clinical Evaluation

App developers are advised to follow the regulations and guidelines listed in Textbox 3.
When Must a Clinical Investigation Be Undertaken?
A clinical investigation is always mandatory for class III devices, regardless of the amount of information that can be retrieved from other sources (Article 61[4] MDR).

Depending on clinical claims, the outcome of risk management, and the results of the clinical evaluation, clinical investigations may also have to be performed for nonimplantable medical devices classified as I, IIa, and IIb. In addition, a clinical investigation must be conducted if there was no sufficient pre-existing clinical investigation data or scientific literature on which to base a clinical evaluation (Article 61 MDR).

Clinical investigations are only to be performed when the information necessary on device performance, safety, and clinical benefit cannot be obtained in any way other than by testing the device on humans (Articles 61,62 MDR).

Notified Bodies
Notified bodies must be involved in the conformity assessment procedures for device classes Iₙ, Iₙₙ, Iₙₚ, IIa, Ib and III (Articles 52,53 MDR). A notified body is an organization designated by national authorities to assess the conformity of certain products with the appendices of the MDR, harmonized standards, and common specifications before being placed on the market. Manufacturers can freely choose between notified bodies that have an expertise in the relevant product area [9]. An official list of notified bodies can be found on the European Commission’s website [10].

What are Common Specifications?
The MDR introduces the concept of “Common Specifications” (CS), which are similar to the already existing harmonized standards (Figure 2). In cases where there are no applicable harmonized standards, insufficient harmonized standards, or where there is a need to address public health concerns, common specifications must be followed in order to demonstrate device compliance with the requirements set by the MDR (Article 9 MDR). One should be aware of this concept since according to Article 9 those CS must be followed. Even when there are no CS available, it is important to stay up to date on developments.

Software Life Cycle

Unique Device Identifier
The Unique Device Identifier (UDI) system was first introduced by the US Food and Drug Administration (FDA) [12] and will also apply to the EU market once the MDR is in force (Article 27 and Annex VI MDR). The UDI-number is linked to the European database of medical devices (EUDAMED), contains all relevant information about a medical device, and is used to identify every device (Articles 27 and 33 MDR). This system has been developed to help track each device, react quickly in case of a serious incident, and prevent marketing of illegal devices (Introduction [41] MDR). This should improve vigilance and consequently patient safety [13]. The UDI code must be affixed as a 2D/Data matrix code, inter alia, ID/linear barcode or radio-frequency identification (RFID), and in a human readable interpretation (HRI) format. The manufacturer must place the UDI on every single product. In case of medical device software and apps, the UDI must be stated within the software, such as in the “about” file or the start-up screen. A medical device for clinical investigation must not have a UDI (Annex VI MDR). According to Article 27 of the MDR, the manufacturer shall keep an updated list of all UDIs as part of the technical documentation. The UDI will be assigned by organizations established for this purpose (Article 27 MDR). Further information can be found in Articles 27-28 and Annex VI (Part C) of the MDR. A guidance document on how to create the UDI database and which data format should be used has been provided by the EU UDI Work Group [14].
Basic-UDI-DI
The Basic UDI-DI is the primary identifier of a device model in EUDAMED and must be referenced in relevant certificates and the EU declaration of conformity. This number will not be affixed to the product. Similar products with the same purpose, such as those that only differ in user interface language, will carry the same Basic-UDI-DI (Part C of Annex VI MDR).

UDI-DI
The UDI-DI is the identifier specific to a device and manufacturer. Software with different user interface languages must carry different UDI-DIs. This part of the UDI is fixed (Part C of Annex VI MDR).

UDI-PI
The UDI-PI is the product identifier used to mark a production series of a device (e.g., batch, serial number, software identification) and is affixed to every single product. Each software version has its own UDI-PI. This part of the UDI is variable (Part C of Annex VI MDR). For software, not every single installation/download will have its own UDI-PI, but each version of the software does require it.

When is a New UDI-DI Needed?
A new UDI-DI is needed in the cases presented in Textbox 4.

Textbox 4. Situations in which a new UDI-DI is needed for apps or medical device software (Part C of Annex VI MDR).
- Changes in performance, efficacy, safety, intended use of the software or interpretation of data
- Changes in algorithms, database structures, operating platforms, architecture, user interface or channels for interoperability
- Change of the software name
- Change in user interface language

When is a new UDI-PI needed?
A new UDI-PI is necessary after minor software revisions such as bug fixes, usability enhancements (those that are not for safety purposes), security patches, or operating efficiency (Part C of Annex VI MDR).

Conclusion
While the implementation of the new, more stringent MDR might lead to the development of more high-quality apps and improved patient safety, it might also limit the development and release of new apps and software on the market. The classification of a device as class IIa or higher requires evaluation by a notified body, which can be very costly and therefore a barrier to entry for app developers. Whether the new MDR and especially rule 11 is a blessing or a curse for app developers will depend on authorities’ interpretation of the guidelines and can only be evaluated after implementation in May 2020.

Authors’ Contributions
LK and USHS conducted the analysis and interpretation of the regulatory framework for medical device development and wrote and revised the manuscript.

Conflicts of Interest
None declared.

References


Abbreviations

CE: Communauté européenne
CS: Common Specifications
EEA: European Economic Area
EMA: European Medicines Agency
EU: European Union
EUDAMED: European Database on Medical Devices
FDA: US Food and Drug Administration
HRI: Human Readable Interpretation
ISO: International Organization for Standardization
MDCG: Medical Device Coordination Group
MDD: Medical Device Directive 93/42/EEC
MDR: Medical Device Regulation (EU) 2017/745
MDSW: Medical Device Software
PIP: Poly Implant Prothèse
QMS: Quality Management System
RFID: radio-frequency identification
TD: technical documentation
UDI: Unique Device Identifier
UDI-DI: Unique Device Identifier – Device Identifier
UDI-PI: Unique Device Identifier – Product Identifier

©Lina Keutzer, Ulrika SH Simonsson. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 26.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Mobile Health for Pediatric Weight Management: Systematic Scoping Review

Louise Tully¹, BSc, MSc; Amanda Burlis², BA, MSc, MBBS, FPH; Jan Sorensen³, MSc, MA; Riyad El-Moslemany⁴, BA, MSc; Grace O’Malley¹,⁵, BSc, MSc, PhD

¹School of Physiotherapy, Division of Population Health Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland
²School of Health Sciences, Division of Health Services Research and Management, City University of London, London, United Kingdom
³Healthcare Outcomes Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland
⁴Department of Psychology, Royal College of Surgeons in Ireland, Dublin, Ireland
⁵W82GO Child and Adolescent Weight Management Service, Children’s Health Ireland at Temple Street, Dublin, Ireland

Abstract

Background: The prevalence and consequences of obesity among children and adolescents remain a leading global public health concern, and evidence-based, multidisciplinary lifestyle interventions are the cornerstone of treatment. Mobile electronic devices are widely used across socioeconomic categories and may provide a means of extending the reach and efficiency of health care interventions.

Objective: We aimed to synthesize the evidence regarding mobile health (mHealth) for the treatment of childhood overweight and obesity to map the breadth and nature of the literature in this field and describe the characteristics of published studies.

Methods: We conducted a systematic scoping review in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews, by searching nine academic databases in addition to gray literature for studies describing acceptability, usability, feasibility, effectiveness, adherence, or cost-effectiveness of interventions assessing mHealth for childhood obesity treatment. We also hand searched the reference lists of relevant articles. Studies aimed at the prevention of overweight or obesity were excluded, as were studies in which mHealth was not the primary mode of treatment delivery for at least one study arm or was not independently assessed. A random portion of all abstracts and full texts was double screened by a second reviewer to ensure consistency. Data were charted according to study characteristics, including design, participants, intervention content, behavior change theory (BCT) underpinning the study, mode of delivery, and outcomes measured.

Results: We identified 42 eligible studies assessing acceptability (n=7), usability (n=2), feasibility or pilot studies (n=15), treatment effect (n=17), and fidelity (n=1). Change in BMI z-scores or percentiles was most commonly measured, among a variety of dietary, physical activity, psychological, and usability or acceptability measures. SMS, mobile apps, and wearable devices made up the majority of mobile interventions, and 69% (29/42) of the studies specified a BCT used.

Conclusions: Pediatric weight management using mHealth is an emerging field, with most work to date aimed at developing and piloting such interventions. Few large trials are published, and these are heterogeneous in nature and rarely reported according to the Consolidated Standards of Reporting Trials for eHealth guidelines. There is an evidence gap in the cost-effectiveness analyses of such studies.

(JMIR Mhealth Uhealth 2020;8(6):e16214) doi:10.2196/16214
Introduction

Background
Childhood obesity remains a leading global public health concern, particularly its prevalence among children because of the short-term comorbidities and long-term impacts on psychological well-being, physical development, risk of noncommunicable disease, progression of comorbidities, and the subsequent economic implications [1]. Although the level of fat accumulation in the body is difficult to measure, widely accepted proxy methods for classifying the level of adiposity, such as age- and gender-adjusted BMI centile curves, are available from the International Obesity Taskforce [2,3], as well as the World Health Organization (WHO) [4] and the Centers for Disease Control and Prevention [5]. During childhood and adolescence, the cornerstone of treatment for overweight and obesity is lifestyle interventions, and the evidence shows behavior change techniques, such as goal setting, incentives, family support, and self-monitoring, alongside dietetic support and increased physical activity to be effective [6-8]. For those with more severe obesity, pharmacotherapy and bariatric surgery may need to be considered [9]. For the remainder of this paper, we discuss lifestyle interventions only when referring to treatment.

The complexity of obesity continues to unfold as researchers and practitioners strive to develop both prevention and treatment options that are effective and sustainable [10]. In doing so, many researchers and practitioners have sought to utilize information and communication technology (ICT) and, in particular, the ubiquity of mobile technology in both developed and developing countries, to deliver treatment with wide reach and efficiency [11]. The WHO Global Observatory for Electronic Health defines mobile health (mHealth) as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [12].

Related Work
Robust evidence syntheses have shown mHealth interventions to be effective tools in enhancing care for the management of certain chronic diseases, including asthma and diabetes [13,14], and it is important to assess their potential in other populations with chronic diseases. Ambulatory care via mobile devices for pediatric obesity could provide treatment to those less likely to access services because of practicality and geography, or indeed to those who live chaotic lives wherein the capacity to attend clinic appointments is compromised. Social distancing measures currently in place in response to a global pandemic have also introduced an urgent need for alternative options for outpatient consultations. Utilizing ICT also allows for a means of accessing therapeutic care that may help to overcome issues associated with stigma and obesity. There is also the potential for saving considerable staff and patient time, improving health-monitoring data collection quality and consistency, and allowing for increased self-efficacy on the part of the patient in the management of conditions [14]. In addition to the interest in digital health from the health care sector itself, mHealth in the context of childhood obesity has recently become the focus of commercial interest also [15], which further reinforces the need for a review of the published evidence.

Previous use of telehealth for the treatment of childhood obesity has shown to be promising, particularly for reaching rural and less-accessible patients [16], and carefully designed mHealth interventions have the potential for improving this reach, given the increasing popularity of mobile electronic devices. The widespread use of mobile electronic devices, in particular, smart devices (such as mobile phones and tablet PCs), has accelerated in the last decade leading to two-thirds of the world’s population being connected to mobile devices [17]. Further, practitioners involved in pediatric weight management have demonstrated openness to the use of mHealth to support treatment [18].

With this rapid social transition to the use of handheld and mobile technology within all aspects of daily life, there has been a sharp increase in research that incorporates mHealth [19]. Despite this increase, challenges remain with respect to augmenting, complementing, or even substituting face-to-face treatment of overweight and obesity with technology. De Jongh et al [13] highlighted the need for further assessment of long-term effects, acceptability, costs, and risks of mHealth interventions. The promotion of mobile devices for health care in this population (eg, children aged <12 years) could be viewed as contrary to the WHO [20] and local guidance to minimize screen time for children, and this is a potential source of confusion and perhaps adverse effects.

There is also the consideration of whether transferring face-to-face clinical services to platforms, which rely on considerably expensive devices could negatively impact on existing health inequalities. Ownership of mobile devices is widespread across the various socioeconomic categories; however, it is still possible that this could further isolate the most vulnerable groups living in poverty, a well-documented driver for obesity. Researchers must also be mindful of digital literacy issues and their impact on inequalities. Previous research has suggested that children with the lowest socioeconomic status are likely to benefit the least from obesity prevention interventions [21], and this should also be considered carefully in relation to treatment efforts.

Many potential advantages of using mHealth also need to be balanced with data protection and privacy considerations and protocols, which vary globally [22]. There may be an unintended risk of compromising children’s privacy and safety on the Web by enrolling them in mHealth interventions, particularly if ownership or access to data is not specifically detailed. Potential adverse events concerning physical safety due to distractions from the environment because of mobile phones is also a consideration for those interested in pediatric mHealth interventions.
In addition, those seeking to leverage mobile technology for the improvement of health care should do so while also assessing its cost-effectiveness. The financial resources required to establish, maintain, and future-proof mHealth interventions may be easy to overlook compared with the more obvious and visible resources required to run a physical in-person obesity clinic. The development and maintenance of modern ICT is associated with significant economic, environmental, and ethical costs. The global demand for ICT (data centers, networks, and connected devices) corresponds to substantial global carbon emissions, and its true cost is difficult to measure [23]. Moreover, the sourcing of materials necessary for manufacturing mobile devices are the subject of ethical concern [24]. These fundamental environmental and social costs are in addition to the cost of translating lifestyle interventions to Web-based or mobile platforms, costs related to design, development, and delivery of software, and cost of testing efficacy in robust trials and rolling out interventions (delivery, evaluation, and maintenance). The true impact of seemingly cost-effective alternatives to conventional health care may be substantial.

Objectives

To date, despite a number of reviews aimed at assessing mHealth for health-promoting behaviors related to the prevention of childhood obesity [11,25], there has not been a review that focused on mobile technology for clinical pediatric weight management. Therefore, we sought to assess what has been researched by mapping the published work and gray literature describing studies of mobile technology for pediatric weight management. We specifically sought to map the methods used and characteristics of studies to present a broad overview of the parameters of work in this field to date and inform future studies that may aim to synthesize findings related to particular outcomes of interest.

This study aimed to assess the breadth and nature of the available literature describing evaluations of interventions using mHealth for the treatment of childhood overweight or obesity.

Table 1. Search strategy.

<table>
<thead>
<tr>
<th>Component number</th>
<th>Search string</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>microcomputer OR telemedicine OR personal digital assistant OR digital health OR wireless OR smartphone OR ‘cell phone’ OR ‘mobile phone’ OR handheld OR mhealth OR app OR OR tablet OR OR PC OR iPad OR OR messaging OR messages OR eHealth OR ‘electronic health’ OR telehealth OR connected health OR internet OR (mobile AND app) OR (mobile AND electronic AND device) OR (mobile AND health) OR (mobile AND application)</td>
</tr>
<tr>
<td>2</td>
<td>child* OR adolescent* OR teen* OR youth OR minors OR pediatric OR paediatric OR young</td>
</tr>
<tr>
<td>3</td>
<td>obesity OR obese OR overweight OR weight loss OR BMI OR “body mass index” OR “body weight” OR “weight management”</td>
</tr>
</tbody>
</table>

The research team carried out two rounds of screening; the first author screened all titles, abstracts and full texts, and a 20.00% (944/4719) portion of titles and abstracts and 25.1% (80/318) full texts were double screened by other reviewers to ensure consistency. Where there were discrepancies between decisions for each stage, both reviewers discussed their decisions and a third researcher was available if needed for consensus.

Inclusion and Exclusion Criteria

Studies were eligible for inclusion if they assessed lifestyle interventions using mHealth for weight management: lifestyle treatment aimed at reducing adiposity or related clinical measures or maintaining weight following treatment) in children and adolescents ≤18 years with overweight or obesity (as defined by local criteria). We included studies if the mobile component

Methods

Design

We used a scoping review methodology [26-29] to achieve the research aim, with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for scoping reviews (PRISMA-ScR) as guidance throughout the reporting process [28]. Our objective was to provide a descriptive map of the characteristics of studies to date in the broad context of mHealth for pediatric weight management, in line with Arksey and O’Malley’s [26] framework for scoping reviews. We sought to include evaluations of mHealth interventions, in addition to studies assessing feasibility, acceptability, adherence, or cost. A protocol for this study was registered on Open Science Framework [30].

Study Selection

We constructed the search strategy (see Multimedia Appendix 1 for full strategy) iteratively with guidance from a research librarian, adapting this where necessary for the various search engines. There were three components of the search strategy, which incorporated (1) search terms for all related terms to the technology and potential devices used, (2) all search terms which might characterize children (ie, anyone aged ≤18 years), and (3) all terms identified related to obesity and weight management (Table 1).

We undertook systematic searches of academic databases, including PubMed, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, Health Business Elite, Medical Literature Analysis and Retrieval System Online, Excerpta Medica dataBASE, Cochrane, Emerald, Scopus, and Web of Science, for articles published between January 2000 and December 2018 to ensure the relevant technology was included. In addition, we searched for gray literature using Connecting Repositories [31], OpenGray, Rian and Bielefeld Academic Search Engine, and also by hand searching the reference lists of relevant articles. Articles were limited to those published in English. We included gray literature such as reports, theses, and conference proceedings for completeness (if the same results were not included as a full article) [32].
was the primary mode of intervention delivery for at least one study phase, or if the mobile component was independently assessed. Qualitative, quantitative, or mixed and multiple method studies assessing acceptability, usability, feasibility, effectiveness, cost-effectiveness, fidelity, or adherence were eligible to achieve our aim of mapping the breadth and nature of the literature in this field.

We excluded studies aimed at preventing overweight in children of normal weight. We excluded studies in which participants were inpatients, or the intervention was aimed at managing underweight participants. Interventions in which the primary purpose of the mobile technology component was collecting outcome data (e.g., a smartwatch to collect data on physical activity) were only included if the intervention also involved tailored feedback or counselling based on the data collected from the device. Studies in which the digital component comprised Web 2.0 platforms that are commonly accessed via apps, but can also be accessed using computers, were only included if the authors specified that only a mobile device would be used by the participants. If it was unclear what type of device would be used, then studies were excluded. The full inclusion and exclusion table used can be found in Multimedia Appendix 2.

### Data-Charting Process

Assessment of eligibility was aided by a decision tool developed for this study (Multimedia Appendix 3), and studies were then categorized according to the characteristics agreed upon a priori [30], using a data extraction form developed for this review. The characteristics were as follows: aim, design, participants, nature of mHealth, the outcomes assessed and measures used, and details of the behavior change theory (BCT, if any) underpinning the mHealth intervention. Data were extracted by the first author using the predefined form.

### Results

#### Study Selection

We identified 8804 titles through database and gray literature searching (Figure 1), and upon removal of duplicates and other ineligible data sources, we screened 4718 titles and abstracts for eligibility with 318 full texts screened thereafter. The initial agreement between reviewers during the title and abstract screening phase was 86.1% (813/944), and 87.5% (70/80) for the full-text screening. All conflicts were resolved through discussion between the reviewers involved in each stage. In both phases, initial disagreement had been because of over-inclusiveness on the part of the second reviewer, which on a more detailed discussion in the context of the eligibility criteria resulted in complete agreement.
Nature of the Literature

We identified 42 studies (based on 25 interventions), which met our eligibility criteria (Multimedia Appendix 4) [33-77]. Of our included studies, 32 were published journal articles and 9 were conference proceedings; 1 doctoral thesis was also included. Of the studies included, 7 were aimed at assessing acceptability among participants [33-38,78], and 2 were usability studies [39,40]. In all, 15 were either feasibility [41-45] or pilot [46-55] studies, and 17 reported outcomes of trials or field studies [35,56-72]. We also identified 1 process evaluation [73]. We did not identify any economic evaluations of mHealth for childhood obesity treatment.

Notably, we identified 30 additional research protocols or registered trials (not included in our review) for studies assessing mHealth for childhood obesity treatment, which were ongoing or not yet published. The majority of these (63%, 19/30) are based on mobile apps as the primary mode of delivery, with 17% (5/30) incorporating wearable technology, and 13% (4/30) using SMS as the only mHealth component.

Participants

The sample size of included studies ranged from 3 to 262 participants. The majority of studies included a small sample size; 45% (19/42) studies included <25 participants, while 67% (28/42) studies included <50 participants. Overall, 29% (12/42) studies included more than 100 participants, and these were all randomized controlled trials (RCTs) (Multimedia Appendix 4), although 3 were pilot or feasibility trials [43,49,55].

All of the interventions aimed at treating childhood obesity and, therefore, included children or adolescents with overweight or obesity. The precise criteria used for the participant inclusion were not always specified, but where they were specified, these were predominantly based on BMI centiles for age and gender, and varied from ≥85th centile to ≥98th centile (see Multimedia Appendix 4). Tripicchio et al [66] included the widest age range and the youngest sample of children, with participants aged 2 to 18 years in a family-based intervention, while Kim et al [38] included the oldest participants, with an age range 13 to 29 years. There were also 3 studies where the intervention focused on the parents [37,59,69]. One additional study aimed at young people with intellectual or developmental disabilities (IDD) [46] targeted parents for a qualitative acceptability study to assess their preferences [78]; however, the main intervention was tested with children and young people [46]. Aside from the studies by Ptomey et al [46,78] for children with IDD, only 1 study had additional inclusion criteria to BMI classification; Patrick et al [60] also specified that participants should have two risk factors for type 2 diabetes mellitus in addition to BMI.
≥85th centile. The participant characteristics for each study are also presented in Multimedia Appendix 4.

**Intervention Content**

Of all the studies identified, 69% (29/42) specified a BCT, or a component of BCT, on which the intervention being assessed was based (Multimedia Appendix 4). In terms of content, almost all of the interventions were multicomponent and focused on various aspects of lifestyle treatment for obesity (diet and physical activity predominantly), incorporating food or physical activity diaries, games, encouragement or feedback related to adherence to physical activity and nutrition goals, or general motivation. Kulendran et al [49] assessed SMS for weight maintenance by comparing commitment-based techniques with information only.

In the study by Saez et al [43], the SMS intervention was solely aimed at motivating participants to attend face-to-face sessions. One study focused only on diet [47], and 2 on only physical activity [54,67]. Similar to the study by Saez et al [43], the SMS component described by Herget et al [67] was primarily aimed at encouraging attendance at physical activity sessions. Overall, 5 of the included studies assessed interventions based only on the self-monitoring aspect of lifestyle treatment, with the technology aimed at recording diet and physical activity [35,38,45,50,61].

**Modes of Delivery for Mobile Health**

The two predominant modes of intervention delivery via mHealth were SMS text messaging and mobile apps. The earliest studies identified were published in 2010 [41,42,62], and SMS remained the most studied form of mHealth for treatment in this population until 2014, at which time, apps subsequently overtook SMS in frequency reported in the published literature. Figure 2 outlines the overall number of studies identified by year and mode of delivery. SMS remained a popular component of evaluated mHealth interventions after 2015 but began to feature as secondary to other forms of mHealth tools including apps and wearable technology [51,61,68]. A total of 6 studies featured wearable technology as the primary mHealth component. Wearable technology goes hand in hand with mobile apps, which are often used for monitoring and collecting the data. Each of these 6 studies also featured at least one app. In all, 3 studies investigated the modes of mHealth other than SMS, apps, and wearables. In 2012, Woolford et al [34] explored an intervention based on Photovoice, where participants used picture messaging as part of the intervention, while Oliver et al [45] explored a novel method of self-monitoring using a personal digital assistant.

**Figure 2.** Number of eligible studies by year and mode of delivery.

**Outcomes Measured**

**Adiposity-Related and Cardiometabolic Outcomes**

For the studies assessing the effectiveness of interventions, the most commonly measured outcome was change in BMI z-score (alternatively referred to as standard deviation score) or BMI percentile using anthropometric measures, which 43% (18/42) of the included studies assessed (Multimedia Appendix 4). Herget et al [67] measured skinfold thickness too, whereas 4 studies reported additional clinical outcomes including blood pressure, biochemical samples, physical fitness, and insulin resistance [63-65,69,72].

**Dietary Measures**

In all, 31% (13/42) studies reported outcomes related to dietary intake or eating behavior; however, the measures used to assess these varied substantially, with nine specified and two unspecified outcome measures. The measures used included the Dutch Eating Behavior Questionnaire [44,62,79], 24-hour dietary food records [54], 3-day food records [58].
photo-assisted 3-day food records, in addition to the Healthy Eating Index 2010 [46,80], food diaries [61], food frequency questionnaires [63,64,69], and items from the California Health Interview Survey [51,68,81]. Durrer et al [72] (a conference abstract) reported using sequential photogrammetry to measure eating speed. The study by Durrer et al [72] and another conference abstract by the same research team [53] also reported using validated questionnaires to assess eating disorders but failed to specify the exact measures. Finally, Pretlow et al [48] assessed (a) whether participants could identify “problem foods” and withdraw from them, (b) whether participants were able to eliminate snacking; and (c) the extent to which participants were able to reduce the amounts of foods consumed at home meals as part of their implementation of an addiction model, which was tracked using the mobile app.

**Physical Activity**

Overall, 4 of the studies included in this review assessed physical activity, objectively measured using accelerometers or wearable technology, as an outcome [46,54,58,72]. One study assessed physical activity by measuring engagement with a fitness app [66] and another by attendance at a high-intensity interval training program [67] (complemented by a variety of self-reported questionnaires). Additional self-reported measures included items on activity habits from the Youth Risk Behavior Surveillance Study [69] and physical activity and sedentary behavior items from the California Health Interview Survey [51,68].

**Psychological Outcomes**

A total of 6 studies that measured psychological health used a health-related quality of life (HRQoL) scale. The measures of HRQoL included the Child Health Questionnaire—PF50 [56], the Pediatric Quality of Life InventoryTM [60,65], the Perceived Quality of Life Scale Adolescents [51,68] and the KIDSCREEN 27 [67]. Further outcome measures related to psychological health and well-being included the Rosenberg Self-Esteem Scale [60], the Mental Health Inventory, the MacArthur Scale of Subjective Social Status, Harter Self Perception Profile for Adolescents, and sex-specific body dissatisfaction scales [63,64]. Chen et al [51,68] assessed self-efficacy using items from the Health Behavior Questionnaire, and Armstrong et al [69] (parent self-efficacy) using the Global Self-Efficacy scale, whereas Herget et al [67] measured self-efficacy, internalization of stigmatization, and perceived social support using validated questionnaires specified in the article. In addition to HRQoL, de Niet [56] measured self-perception using the Dutch version of the Self-Perception Profile for Children, which measures self-perceived competence. Kowatsch et al [70] measured the emotional and social relationship between the participants and their chatbots, using a short version of the attachment bond scale of the Working Alliance Inventory. The conference abstracts by Durrer et al [72] and Lallemand et al [53] reported the assessment of mental health, mood, and well-being as well as motivation, but the measure or measures used were not stated. Finally, Pretlow et al [48] also measured addiction guilt, stress, control, and self-esteem using individual self-report items.

**Process Outcomes**

We identified 3 studies which formally assessed the usability of the mHealth intervention. Oliver et al [45] used an adapted version of the System Usability Scale by Brooke [82] to assess their electronic dietary record, while Ptomey et al [46] administered Likert-scale questions on participant comfort using the tablet and its various features relevant to their specific intervention. O’Malley et al [39] measured technical effectiveness, technical efficiency, and usability via the Software Usability Measurement Inventory (SUMI) [83], which measures efficiency, effect, helpfulness, controllability, and learnability.

Acceptability was also widely measured, with 15 studies reporting patient experience of, or satisfaction with, the intervention. For the most part, this was assessed quantitatively using surveys, predominantly Likert-scales or similar rating-based survey items specifically designed for the individual studies [38,40,43,48,51,61,66,67]; however, Jensen et al [50] and Nguyen et al [73] used previously validated tools, the Client Satisfaction Questionnaire [84] and an adapted version of a satisfaction questionnaire by Golley et al [85]. In all, 8 studies assessed the acceptability of the intervention using qualitative methods [33-37,42,44,78], whereas an additional 5 studies included open-ended questions or interviews to supplement quantitative assessment of acceptability by collecting additional feedback [38,40,43,50,66].

Although adherence was the primary outcome for just 1 study [57], a further 12 studies reported adherence with the intervention as an outcome. Adherence was predominantly measured using the data for direct engagement with the technology [46,47,66,70], as well as responses to SMS communication [41,55,57,59,62], attendance at the face-to-face sessions [57,69], and the level of self-monitoring [25,50]. Nguyen et al [73] reported facilitator adherence to the program protocols (ie, fidelity) as well as participant engagement with the intervention as a whole; however, these were not specific to the mHealth component (SMS).

**Discussion**

**Principal Findings**

This review presents an overview of the literature on mHealth for pediatric weight management and its characteristics. Our study highlights substantial heterogeneity in the interventions, designs, participants, and outcomes assessed in studies that have evaluated the use of mobile technology for the treatment of childhood overweight and obesity. Our findings show that the majority of the work thus far has been aimed at assessing feasibility, acceptability, or usability of mHealth interventions for lifestyle treatment, with these forming parts of an emerging evidence base.

There is no doubt that current technology is rapidly developing, and this poses a difficult challenge for researchers to design and test technology using robust experimental methods while remaining relevant. This review reports that until recently, interventions based solely on text messaging comprised the extent of mobile therapeutic care in this population; however, there has been a sharp transition to apps, often complemented
by wearable technology in the last 4 to 5 years. SMS remains a part of many interventions but is no longer necessarily considered the focus of research as is often the case with technological advances which have become embedded in practice [86]. Although our review shows that much of the recent work carried out was largely aimed at perfecting the technology and ensuring its feasibility in this population, it is clear that we can expect a continued rise in emerging evidence for its effectiveness in the near future as these studies give way to full trials. However, there is a possibility of marketing of mHealth interventions by commercial entities as treatments before completing the testing through experimental methods with the target end users. We identified only 1 systematic review [87] published in 2015, with only two included studies, that specifically met our inclusion criteria, indicating that this was until recently, a field in its infancy.

We investigated the literature specifically aimed at the treatment of childhood overweight and obesity using mHealth, which eliminated a large number of studies that also focused on prevention in this age group or included a high proportion of participants with normal weight. Many of the published systematic reviews have either included mHealth interventions for both prevention and treatment [88-90], or interventions aimed at specific behaviors associated with obesity (eg, sedentary behavior) [91,92]. However, owing to the complexity of overweight and obesity and the specific needs and challenges associated with treating rather than preventing excessive weight gain, we chose not to combine the two. Children with clinical obesity are a specific population where the accumulation of excess adipose tissue may already be affecting the body’s structures and functions and the child’s health and well-being. As such, obesity was defined as a disease by the WHO in the 1970s—a concept revisited more recently by the Childhood Task Force of the European Association for the Study of Obesity [93]. Given that obesity treatment involves a child or adolescent presenting to the health care system for support or help with this condition and related comorbidities, the type of interventions that may be effective are likely very different to that offered to a child for the prevention of obesity. In addition, clinical health services are more commonly required to address the treatment of obesity rather than its prevention and as such, we excluded prevention studies.

There is evidence that while an array of apps aimed at addressing childhood obesity exist, many lack inputs from health professionals with experience in treating childhood obesity or patients living with the condition [94]. A review of nutrition apps relevant to childhood weight management demonstrated that the majority of those available were not evidence based [95]. Our findings suggest that in the academic literature, BCT features prominently in many of the mHealth interventions developed, although in some cases these are components of BCTs such as self-monitoring, which when used in isolation may not be effective. We did not formally critically appraise the studies included in this review given our study design, but future reviews of interventions in this field should aim to do so, in particular, to evaluate the use and appropriateness of BCTs and their application to mHealth for this purpose. In addition, while outcomes including usability and acceptability were frequently measured within the studies we identified, there are no universally accepted definitions for these and the measures used for them were variable, often not validated at all or not validated for children. The authors are aware of ongoing Delphi consultations to develop a common definition and approach to measuring user experiences with pediatric electronic health interventions, which will be extremely useful in designing future interventions for this population.

The heterogeneous nature of mHealth is important to consider for comparing studies in the future. Automated text messaging, for example, is quite dissimilar to interactive apps. Furthermore, some types of mHealth are more closely related to other electronic means of treatment or telemedicine, and it may be more useful to compare the type of communication (eg, educational messages) rather than define these by the mode of communication (mobile device, email, or paper). Thus, limiting reviews to mobile devices means that intervention methods, which are similar but do not necessarily fall into the mHealth category (eg, Skype vs FaceTime counselling, or serious games for smartphones vs games consoles) may be counterproductive. Future reviews should possibly take into account this diverse nature of mHealth and aim to narrow meta-analyses, in particular to interventions which are more closely aligned.

The most common health-related outcome measure across all studies included in this review was BMI z-score or similar. Although change in BMI z-score provides a useful standardized proxy for change in adiposity, it is ultimately flawed as it represents size and shape rather than fat accumulation [96]. This remains a challenge that is universal to the study of childhood obesity. However, for future trials and reviews that assess the effectiveness of mHealth interventions, it is vital that attention is given to the definition of effectiveness, particularly where participants have multimorbidities and small changes in behavior or level of adipose tissue, which while not necessarily resulting in significant change in BMI, may still represent a positive outcome. For children and young people with unique needs and circumstances, such as those with autism spectrum disorder who are at increased risk of overweight and having secondary health issues [97], assessing BMI or indeed using general dietary measures such as food frequency questionnaires may not be useful. The studies in our review that assessed diet and physical activity did so using a very diverse set of measures, particularly for dietary intake. Diet is especially difficult to measure as doing so objectively is not practical and often not feasible and comparing tools with varying levels of accuracy and precision will present a challenge in assessing this as an outcome across studies. None of the measures of diet highlighted in this review are considered to be the most reliable [98]. Moreover, self-reported dietary intake measures have been demonstrated to be less reliable in children with higher BMI compared with those with lower BMI [99] and were developed for use in healthy adult populations. This will be an important consideration for critical appraisal of the evidence in future reviews, in addition to assessing adherence to the Consolidated Standards of Reporting Trials for eHealth (eCONSORT) [100] guidance for Web and mobile trials, which none of the studies included in this review referred to.
We did not identify any cost-effectiveness analyses of childhood obesity treatment via mHealth, which would provide important information to decision makers. Economic evaluations of such studies are needed because mobile technology is frequently cited as a cost-effective [37,101,102] alternative to in-person treatment in this population, in the absence of any actual study or formal assessment of the costs.

Finally, although none of the studies in our review observed adverse events related to the use of the mobile device, future studies will need to report on unintended events in a systematic manner and be mindful of whether mHealth interventions increase the risk of negative effects such as exposure to digital marketing or safety concerns because of distraction (eg, road traffic accidents while using devices) [103].

**Strengths and Limitations**

This study has a number of strengths. Our focus on participants with overweight and obesity specifically, while including a broad range of study designs, has allowed us to provide a comprehensive overview of the literature at all stages of the research process in relation to mHealth for the treatment of childhood overweight or obesity. This is the first review to do so to date. This will provide a useful basis for researchers and health care professionals to identify gaps in the literature and areas for development or to facilitate defining clear questions regarding the effectiveness of interventions on specific outcomes of interest (eg, measures of usability or body composition). There are also some limitations. We did not include Web 2.0 interventions, such as those utilizing Facebook, which participants would often use on a mobile device. As we cannot guarantee the usage on a mobile device, it was not possible to distinguish interventions that were exclusively used on a mobile device from those that might be used on desktop computers. We aimed to minimize the publication bias by including gray literature and conference proceedings; however, it is possible that some conference abstracts were not identified using our search methods given the varying levels of indexing for these. Some bias may also have arisen from our exclusion of studies not published in English. In addition, while a second reviewer thoroughly screened a portion of all titles and abstracts and full texts, the entire set of search results was only screened by one reviewer. The initial large percentage of disagreement in studies to include may, however, represent bias on the part of the reviewers or indeed ambiguity in the inclusion criteria, although the team addressed this using the best available means, through discussion.

**Conclusions**

In summary, the field of mHealth for the treatment of childhood overweight and obesity is new, and the evidence base is still emerging; however, it is certainly a rapidly developing research area. The studies to date have mainly aimed at assessing the feasibility of interventions and are heterogeneous in nature with a diverse variety of outcome measures used. There is a need for cost-effectiveness studies alongside further large, robust RCTs, which employ valid outcome measures and report following eCONSORT guidelines.

**Acknowledgments**

The authors wish to thank Grainne McCabe, Scholarly Communications & Research Support Officer at the Royal College of Surgeons in Ireland, for her support in constructing the search strategy for this study. This work was funded by the Royal College of Surgeons in Ireland, StAR program, and The Temple Street Foundation (PAC11-58). The study was conducted as part of the Health Research Board (HRB) SPHeRE Program (Grant No. SPHeRE/2013/1). The HRB supports excellent research that improves people’s health, patient care, and health service delivery.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
PubMed search strategy.
[DOCX File , 13 KB - mhealth_v8i6e16214_app1.docx ]

Multimedia Appendix 2
Full inclusion and exclusion criteria.
[DOCX File , 14 KB - mhealth_v8i6e16214_app2.docx ]

Multimedia Appendix 3
Decision tool for screening.
[DOCX File , 128 KB - mhealth_v8i6e16214_app3.docx ]

Multimedia Appendix 4
Table of included studies.
[DOCX File , 23 KB - mhealth_v8i6e16214_app4.docx ]
References


81. Brook J. Hell - Jens Oliver Meiert. SUS-A Quick and Dirty Usability Scale URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2020-02-05]


Abbreviations

BCT: behavior change theory
eCONSORT: Consolidated Standards of Reporting Trials for eHealth
HRB: Health Research Board
HRQoL: health-related quality of life
ICT: information and communication technology
Overview of Policies, Guidelines, and Standards for Active Assisted Living Data Exchange: Thematic Analysis

Laura X Fadrique¹*, MSc; Dia Rahman¹, BE; Hélène Vaillancourt², PEng, MScA, PhD; Paul Boissonneault³, PEng, MEng; Tania Donovska², PMP, MEng; Plinio P Morita¹,4,5,6,7*, PEng, MSc, PhD

¹School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada
²CSA Group, Toronto, ON, Canada
³Ontario Health (East), Ottawa, ON, Canada
⁴Research Institute for Aging, University of Waterloo, Waterloo, ON, Canada
⁵Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
⁶Health Innovation, Techna Institute, University Health Network, Toronto, ON, Canada
⁷Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

* these authors contributed equally

Corresponding Author:
Plinio P Morita, PEng, MSc, PhD
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 5198884567
Email: plinio.morita@uwatertlow.ca

Abstract

Background: A primary concern for governments and health care systems is the rapid growth of the aging population. To provide a better quality of life for the elderly, researchers have explored the use of wearables, sensors, actuators, and mobile health technologies. The term AAL can be referred to as active assisted living or ambient assisted living, with both sometimes used interchangeably. AAL technologies describes systems designed to improve the quality of life, aid in independence, and create healthier lifestyles for those who need assistance at any stage of their lives.

Objective: The aim of this study was to understand the standards and policy guidelines that companies use in the creation of AAL technologies and to highlight the gap between available technologies, standards, and policies and what should be available for use.

Methods: A literature review was conducted to identify critical standards and frameworks related to AAL. Interviews with 15 different stakeholders across Canada were carried out to complement this review. The results from interviews were coded using a thematic analysis and then presented in two workshops about standards, policies, and governance to identify future steps and opportunities regarding AAL.

Results: Our study showed that the base technology, standards, and policies necessary for the creation of AAL technology are not the primary problem causing disparity between existing and accessible technologies; instead nontechnical issues and integration between existing technologies present the most significant issue. A total of five themes have been identified for further analysis: (1) end user and purpose; (2) accessibility; (3) interoperability; (4) data sharing; and (5) privacy and security.

Conclusions: Interoperability is currently the biggest challenge for the future of data sharing related to AAL technology. Additionally, the majority of stakeholders consider privacy and security to be the main concerns related to data sharing in the AAL scope. Further research is necessary to explore each identified gap in detail.

(JMIR Mhealth Uhealth 2020;8(6):e15923) doi:10.2196/15923
KEYWORDS
ambient assisted living; active assisted living; AAL; Internet of Things; aging well; aging in place; elderly; geriatrics; standards; policies; health care; ambient intelligence; domotics; ubiquitous health; wearable

Introduction

Background

People are living longer than ever before. According to Statistics Canada’s 2016 census, seniors outnumber children aged under 15 years for the first time [1]. The aging population is growing faster than the working-age population. This adds stress on population growth, potential loss of economic productivity, and output, and a significant portion of the gross domestic product is spent on health care and pensions [2]. Aging also brings specific challenges around declining health, increase of chronic diseases, increased need for daily care and monitoring, and the financial burden of expanding health care costs. As a result, health and independence are top priorities for seniors in Canada. Active assisted living (AAL) offers technology that may address some individual and governmental concerns. In Canada, AAL technologies have been explored as a combination of smart-home, telehealth, and assistive technology (AT) [3]. In contrast, Europe has the most advanced programs for AAL standardization, and many of their projects are developed by the AAL Joint National Program [4] to support their growing aging population. Moreover, while AAL has clear applications among an aging population, similar needs exist within vulnerable populations for timely health care, monitoring, and support.

To provide better living conditions and assistance in daily routines for older adults and vulnerable populations inside and outside of their homes, researchers and innovators have explored the use of different technologies such as wearables [5], camera-based sensors [6], and mobile health technologies [7]. These new technologies and products for daily assistance allow individuals to lead a more independent life [8].

Challenges

As mentioned by Rashidi and Mihailidis [9] in their article, the majority of older adults prefer to stay in their homes instead of transferring to a home care facility, and therefore it is essential to create and develop new technologies that support graceful aging in place. However, vulnerable populations have unique needs and limitations. Therefore, it is necessary to understand the differences and challenges in designing and developing technologies as well as collecting data for this specific user group. Standards, guidelines, and policies play an important role, ensuring better quality and safety of new technologies. In addition, standards enable knowledge sharing and act as a mechanism to make appropriately protected knowledge public and widely accessible [10].

The AAL technology landscape may seem relatively new and can sometimes be confused with the Internet of Things (IoT) as both involve data acquisition from the environment using wireless technologies [11]. IoT is the extension of the internet into physical technologies and everyday objects, which enables the creation of systems that operate over a network, collecting and exchanging data, and acting upon objects in our lives [12].

IoT devices can be used for any purpose other than health care, thus differing from AAL technologies that have the primary purpose of assisting users’ quality of life. Although this new field already has many expanding technologies, AAL lacks a rigorous process of development aimed at specific users. In reviewing existing literature, some key studies on AAL technology, standards, and frameworks have been identified. The majority of the studies in this field focus on the technical specifications of AAL technologies, scenarios reviewing products, and the available tools and solutions [9,13-15]. Other studies focus on existing projects and platforms within the AAL field [2,16]. One particular study highlighted the frameworks, platforms, standards, and quality attributes of AAL technology, and so it was used as the baseline of our literature review [17].

The initial literature review and analysis demonstrated the need for researchers to understand the existing technologies and their applications. There is insufficient information on the challenges encountered in creating new AAL technologies or about ways to connect AAL technology through the use of integrated data for the benefit of the user.

Objectives

This paper presents a review of standards, frameworks, policies, and guidelines in the creation of AAL technologies. In particular, the objective of this paper was to provide an overview of the primary standards and frameworks available as well as to highlight the gaps, challenges, and opportunities existing in the AAL area. As such, this paper did not intend to review currently existing technologies or indicate which ones to use or not. Similarly, the scope excluded any standards and frameworks targeted at medical devices even if the device is collecting personal information and is installed in a home environment to improve the quality of life. For this paper, AAL technology consists of all technology, devices, and wearables connected to the internet that enables data collection and exchange and is used for health care monitoring or to enhance the daily life of individuals.

Active Assisted Living Concepts and Terminology

AAL technology is a subset of a broader concept called ATs, which refers to “any item, piece of equipment, or product system, whether acquired commercially, modified, or loop customized, that is used to increase, maintain, or improve [the] functional capabilities of individuals with disabilities” [18]. In other words, ATs can be any tool or device capable of assisting a person to achieve something that would not be possible without the support of that technology [8]. Wheelchairs, walkers, electronic jar openers, screen readers, hearing aids, and educational software are all examples of ATs in everyday use. Despite being a subset of an overarching concept, AAL is also an umbrella term that describes technologies designed to improve the quality of life, aid in independence, and create healthier lifestyles for those who need assistance at any stage of their lives. AAL involves concepts, products, and services
that combine new technologies and environment to improve the quality of life at all ages [19]. AAL uses information and communication technologies combined with the physical and social setting to provide easy-to-use devices either at home or to support lifestyles outside of the home environment [4]. According to the International Electrotechnical Commission (IEC) Systems Committee on AAL (SyC AAL), this technology supports systems for the elderly in industrialized countries by helping them live healthier lives [20]. The term AAL can be referred to as active assisted living or ambient assisted living as both terms are used interchangeably throughout the literature. This research report will follow the terminology defined by the SyC AAL committee that defines AAL as active assisted living technology [20]. The active assisted living terminology was chosen because it goes beyond the ambient and can be used outside the residence such as a smart walker.

Technology could prove highly beneficial in providing a higher quality of life for an aging society or for anyone who needs additional help to perform their daily tasks and activities inside or outside their homes. To address this matter, the International Medical Informatics Association of the United States approved the creation of a new workgroup on smart homes and AAL in November of 2006 [21].

The AAL environment can integrate multisensors inside or outside of the home to gather data and monitor individuals in their homes [22]. The integration of sensors, embedded in homes, is also known as ambient intelligence (AmI) [23]. AmI applications should be transparent to users while meeting security and privacy requirements [24]. Many current devices, sensors, and health/wellness trackers are capable of collecting and sending information to a caregiver or physician for remote patient monitoring. However, the creation of new solutions for the aging population requires special attention, should not rely on the user’s effort, and needs to consider the cognitive, perceptual, or physical limitations of users [9].

AAL technology devices can be either simple, such as presence sensors, or complex such as a smart wheelchair controlled via eye movements. AAL technologies are not subject to the same rigorous standards and evaluation protocols required for medical devices. Standards are widely used by companies in the planning, development, and production of these products and technologies. Without standards, possible interactions between products could be inconsistent, processes would not be defined and secure, and there would be security and safety-related risks [25]. For the Institute of Electrical and Electronics Engineers (IEEE), standards are “published documents that establish specifications and procedures designed to maximize the reliability of the materials, products, methods, and/or services people use every day” [26]. This project focused on identifying existing standards relevant for AAL technology and explored the existing gaps in terms of standards for supporting the development of AAL technologies, and to do that we interacted with different stakeholders in this space. In addition to standards, it is necessary to explore protocols used within the standards. In this case, protocols are a set of rules or procedures for the way information will be structured and transmitted for electronic devices to send and receive data [27].

Methods

Study Design

This project was planned in three phases. The first phase focused on conducting a literature review to understand what currently exists regarding standards and guidelines for AAL technologies. The second phase interviewed key industry stakeholders to develop a better understanding of the use of standards, the use of data-sharing practices, and the challenges of AAL technology to identify the existing gaps in the development of AAL technologies. The third phase aimed to validate the results of the literature review and interview phases through conducting a workshop. The interview method was chosen because it allows more feedback points to be collected from a single individual [28]. As for the workshops, participants promoted group discussions and elaborated on each other's responses, influencing the direction of the workshop [29]. However, the workshops by themselves have disadvantages around the existence of dominant people that dominate the discussion vs interviews that allow users to voice their opinion with regard to that dominance [30].

It is essential to mention that standards or guidelines related to medical devices and their safety were excluded from the scope of the project as there are existing regulations and standards in place to support the development of these devices.

Literature Review

To meet the objectives of this project, a narrative literature review was performed to understand the existing material regarding technical standards, frameworks, and platforms related to AAL technology. Databases used included Scopus, IEC, IEEE, and PubMed as the primary sources for academic references and standards references. The IEC and IEEE databases were selected from the collection of publications related to engineering and computing standards and guidelines. Scopus and PubMed were used to review publications regarding science, technology, health care systems, and medicine. The academic literature led to an in-depth evaluation of gray literature and websites from AAL governmental programs around the world. Furthermore, results from the academic literature leverage the creation of the questions for the semistructured interview used in the next phase of the project. The technology-oriented standards covered in this research report were driven by the research conducted by Memon et al [17] in the article titled “Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes;” a website from Postscapes [31] called “IoT Standards and Protocols;” and Salman’s [32] paper titled “Networking Protocols and Standards for Internet of Things.”

Interviews

In the second phase of this research project, we used a semistructured interview (see Multimedia Appendix 1) with 12 to 17 questions. The semistructured method uses a list of predetermined questions that guide the interview and may or may not be used according to the course of the conversation and previous answers. This method brings out how the interviewee interprets the topics and problems presented [28].
Over 50 stakeholders in AAL technology, from Canada, were invited to participate in interviews. Stakeholders were selected among four distinct categories: (1) health care providers such as physicians, nurses, and social workers; (2) academics and researchers who represent a large percentage of the stakeholders in the field; (3) industry representatives from well-established corporations to small start-ups working to innovate and find better ways to help people; and (4) health care administrators responsible for decision making in research or acquisition of new technologies. The stakeholder list was formed along with a project advisory panel that identified and suggested the names of experts from across Canada with some interest or involvement with AAL. The stakeholders were then divided into the suggested categories.

A round of invitations were sent to all stakeholders, along with an information letter and a description of the project objectives. A total of 15 invited participants agreed to be interviewed. A date and time were scheduled for each participant. The interviews were conducted over the phone by 2 researchers. Each phone call lasted approximately 60 min and was recorded. After a brief introduction of the project, the respondents were presented with approximately 17 questions (the questions could vary according to previous answers) on four distinct areas: “What is AAL?;” ”Standards;” “Data Sharing;” and ”Main Challenge.”

The interview results were coded using a thematic analysis because it is best suited to identifying topics within verbal or written interviews using semistructured interviews [33,34]. Furthermore, the data were analyzed using the 6-phase approach to a thematic analysis proposed by Braun and Clarke [35]. We identified saturation on our themes, instead of saturation on the data.

Workshops

The literature review and interview phases generated a list of standards, platforms, and frameworks, as well as a list of topics. These results were then presented at a workshop organized to fulfill the predefined purpose of validating the results and leverage new insights into and suggestions on the topics presented. A second round of emails were sent to more than 70 stakeholders from Canada, inviting them to participate in a face-to-face workshop with the possibility of online participation, and a total of 11 participants attended the workshop conducted in April 2018. The workshop was created in an unstructured manner where the project researchers acted as facilitators guiding the session. Participants were given the opportunity to present their ideas related to the topics presented and to challenge other participants’ ideas. The expected outcome was to identify the collective understanding of the topics presented and thus build a common meaning and validate the results.

Future steps and opportunities related to AAL technology were identified as a result of conducting the workshop.

Results

Standards

Considering a variety of possible information technology standards, the literature review showed that standards related to essential technologies, hardware, devices, application programming interfaces, and middleware are well-covered by the existing standards of leading institutes such as IEEE, International Organization for Standardization (ISO), and IEC. On the basis of this information, the identified standards relevant to AAL technology were grouped into the four following categories: (1) design and terminology; (2) communication and transport; (3) privacy and security; and (4) data content. For this study, the design and terminology group was responsible for representing concepts using correct terminology and processes related to design, modeling, and planning. Any standards responsible for ensuring the information were transmitted reliably, and independent of the message sent, they were classified as communication and transport. Privacy and security standards are responsible for setting administrative, physical, and technical actions to protect the confidentiality, availability, and integrity of the information. The data content group contains standards responsible for the transferred information and data format that usually uses existing communication protocols. The frequencies of the different groups are shown in Table 1, where the first column displays the standard group and the second column lists the number of standards or protocols identified for a given group as well as the percentage relative to the total. Figure 1 represents the primary standards and protocols investigated in this report and the existing association between them. The figure is a radial chart, sectioned in four categories. The top-left area (solid yellow line with no fill) shows the standards responsible for data content; the top-right area (dotted blue line with no fill) shows patterns used for design and terminology; the bottom-left area (solid grey line with pattern fill) shows the security and privacy standards; the bottom-right area (solid orange fill) shows patterns related to communication and data transport. The link between the standards can represent a dependency—in this case, one standard does not exist without the other—or the indication that one standard is based on another one.

Table 1. Number of standards for each standard category identified (N=43).

<table>
<thead>
<tr>
<th>Standard group</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design and terminology</td>
<td>5 (12)</td>
</tr>
<tr>
<td>Communication and transport</td>
<td>15 (35)</td>
</tr>
<tr>
<td>Privacy and security</td>
<td>13 (30)</td>
</tr>
<tr>
<td>Data content</td>
<td>10 (23)</td>
</tr>
</tbody>
</table>
Figure 1. Relationship between standards and protocols.

### Framework and Platform

Beyond the standards and protocols, several frameworks and platforms were identified as relevant for the AAL context. These frameworks bring together multiple standards and guidelines, thus enabling products compatible with the platform or framework to integrate with other products of the same family with ease. Some frameworks or platforms were created for health system purposes or specifically for AAL technology, including the following:

1. **Continua framework**: Continua is a framework created by the Personal Connected Health Alliance based on open standards. The framework is known as a guideline for the safe, secure, and reliable exchange of data to and from personal health devices [36].

2. **Persona platform**: Persona platform is an integrated project financed by the European Commission on AAL for an Aging Society to develop sustainable and affordable solutions for senior citizens to live independently [37].

3. **UniversAAL**: UniversAAL is an open-source middleware platform for AAL that enables the rapid development of innovative IoT solutions [38].

4. **ZigBee Healthcare**: ZigBee Healthcare (proprietary) is an AT that is designed to be simple and easy for users to maintain their independence and mobility. This type of network can connect to sensors and controllers without being restricted by distance and range [39].

5. **Microsoft HealthVault**: Microsoft HealthVault is a proprietary platform built by Microsoft to support the collection, storage, use, and sharing of health information for users, family members, and care providers. It allows all health information to be accessed from a single place.

6. **IEC/SyC AAL**: The IEC SyC AAL has developed an international roadmap for standards for AAL systems and
services to ensure safety, security, privacy, and interoperability [20].

7. Apple HealthKit: Apple HealthKit is a proprietary platform available for iPhone users to collect and aggregate health data via wearables and apps that are installed or synchronized with the users’ iPhones or Apple Watch. HealthKit provides a standardized framework for the storage and sharing of health information, allowing users to control their data access and integration [40].

Interviews and Workshops

Interviews have highlighted that approximately 40% of the interviewees were well versed with the term AAL, while 27% had heard of the term but could not explain the meaning of the abbreviation and 33% did not know the terminology. After a brief explanation of the AAL concept, 60% of the interviewees already knew the idea even though they were not familiar with the specific terminology. The interview participants also confirmed that the terminology is considered a problem, with the first issue being the acronym AAL that can mean either ambient assisted living or active assisted living depending on the particular context or group. Another concern is regarding the existing stigma with the term assisted because it implies the need for assistance and support, which several technology users do not desire. These findings were confirmed by the workshop participants as well.

Questions related to standards showed that all the interviewees agree on the use and creation of specific standards and guidelines of AAL technology. Participants in academia pointed out that even though standards are not fully applicable in the research area, they are of paramount importance in the development of new technologies. In addition, standards-related responses showed concerns with end users, user safety, product accessibility, and the purpose of using the technology.

When questioned about data sharing, privacy and security were identified as the major challenge. The participants also expressed concern about proper interoperability between products to ensure the correct exchange of information. However, all participants agreed that they would share their data for research and to improve public health if adequate safety policies were implemented and data anonymity is practiced. Another point raised was related to data accessibility, especially in the context of the elderly and vulnerable populations.

Regarding the challenges of creating new AAL technologies, most participants understood that technology challenges or lack of technology is not the problem in the creation process. If the technology does not exist yet, it will probably be created. The problem lies in ensuring security, privacy, proper data sharing, product interoperability, and the correct use of technology. It is essential to understand the purpose of the product being developed to be able to select appropriate technology to ensure the greatest benefit to the end user.

The Five Challenges

After the round of interviews, the notes were analyzed together with a revision of the standards and framework, and five significant gaps were identified: (1) end user and purpose, (2) accessibility, (3) interoperability, (4) data sharing, and (5) privacy and security. Table 2 shows the number of mentions for each gap by the type of stakeholder. Home care administrators are primarily concerned with the benefits for the end user and if the proposed technology does what it is intended to do, while the industry and health care providers place more value on privacy and security. Furthermore, academics mention data sharing as the biggest challenge. Figure 2 shows the frequency of each gap mentioned during the interviews and workshops. Privacy and security is the primary issue, with 30% of mentions, followed by data sharing with 29%, and end user and purpose with 23%. Accessibility is the least mentioned gap with 5%, and the second-lowest gap is interoperability with 13% of mentions. Although interoperability is at the end of the list, it is identified in the literature to be the most significant technical challenge within AAL technologies.

Table 2. Number of times a gap was mentioned by a group of stakeholders—interview result.

<table>
<thead>
<tr>
<th>Gap group</th>
<th>Accessibility</th>
<th>Interoperability</th>
<th>End user and purpose</th>
<th>Data sharing</th>
<th>Privacy and security</th>
<th>Group total, N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health care providers, n (%)</td>
<td>1 (14)</td>
<td>0 (0)</td>
<td>1 (14)</td>
<td>2 (29)</td>
<td>3 (43)</td>
<td>7</td>
</tr>
<tr>
<td>Academic, n (%)</td>
<td>1 (4)</td>
<td>3 (13)</td>
<td>4 (17)</td>
<td>10 (43)</td>
<td>5 (22)</td>
<td>23</td>
</tr>
<tr>
<td>Industry, n (%)</td>
<td>1 (3)</td>
<td>4 (13)</td>
<td>6 (20)</td>
<td>7 (23)</td>
<td>12 (40)</td>
<td>30</td>
</tr>
<tr>
<td>Home care admins, n (%)</td>
<td>1 (4)</td>
<td>4 (17)</td>
<td>8 (35)</td>
<td>5 (22)</td>
<td>5 (22)</td>
<td>23</td>
</tr>
<tr>
<td>Gap total, n</td>
<td>4</td>
<td>11</td>
<td>19</td>
<td>24</td>
<td>25</td>
<td>83</td>
</tr>
</tbody>
</table>
End User and Purpose

AAL technologies are meant to assist users and keep them and their data safe. There is an overall perception that there is not enough consideration for the end user during product development. The end user should participate in all the phases of development, helping in the planning, designing, and testing of new technologies to ensure that their specific impairments, diseases, and disabilities are being accurately addressed. The technology should be adaptive to the context and satisfy the users' needs. Another concern raised by the participants is that a significant portion of the technology available in the market was not developed to solve a clinical problem. Instead, it was adapted from the original purpose (eg, home security) to solve an alternate problem.

Accessibility

The access to and use of technology, regardless of user ability, is an essential aspect of AAL technology. Accessibility is one of the themes identified by interview and workshop participants as differentiating average users from AAL users. The word accessibility has emerged as one of the top 5 significant gaps related to standards and guidelines within the scope of AAL technology. Creating products that are accessible to all is where the effort must be concentrated to ensure maximum benefits from technology. Lack of accessibility can lead to a decrease in device acceptability due to a deficiency of support for AAL users' needs.

Interoperability

The lack of interoperability between AAL technologies was one of the most predominant technical challenges in the reviewed literature and among the participants interviewed. Owing to the lack of regulation, the manufacturer enables integration with other devices that use the same protocol when choosing a solution (eg, the ZigBee protocol), disabling integration for devices that have opted for other alternatives. Many of the standards, protocols, and framework presented in this paper have the goal of solving this problem.

Data Sharing

Standardizing the data-sharing process between devices is one of the major challenges in the field of AAL technologies. Most devices cannot communicate with each other due to a lack of proper interoperability. Even when they do use the same technology, data exchange is not always feasible. As such, the challenges related to data sharing between AAL devices are related to the availability, reliability, integrity, validity, and accuracy of the collected data. In particular, how to ensure that data collected on one device are transmitted to another device without loss of information and quality. The interviews and workshops reveal that it is necessary to work with the terminology of the data so that the information is significant and creates unified terminology models across all manufacturers. In doing so, it is possible to transform data and present the results to the end user in a clear and understandable way without the need for technical or specialized knowledge for data interpretation.
Privacy and Security

Security concerns range from technical issues—whether devices are protected against viruses or hackers—to ensuring that devices are designed and developed with security in mind by choosing the best algorithms and encryption available. Devices that serve more than one purpose run a higher risk of having security and privacy requirements that are not correctly designed. Ensuring that the data collected are properly anonymized and aggregated where appropriate, so as not to pose a potential risk to the end user, is the most significant privacy concern. Most participants report the need for a transparent process that clarifies how data flow across the internet or devices, which data are being collected, and how the data are used. There should be a clear explanation of when and how data are shared and who has access to the information. Trust will only be achieved with proper end-user education, transparency, and accessible presentation of end-user policy contracts.

Discussion

Overview

This study used literature reviews, interviews, and workshops to identify existing patterns, structures, and guidelines for the development of AAL technology, as well as to identify existing gaps in the area.

Our research has shown that the Canadian aging population could benefit from the innovation of AAL technology in the coming years. AAL technology can provide solutions to increase the security and independence of the population, as well as improve the quality of life, allowing seniors to age in their homes. Therefore, it is crucial to continue investing in projects and solutions to improve the development of AAL technologies.

Interestingly, the results showed that most of the challenges within the scope of this problem are not related to the availability of technology but to the way technology is applied to solve current problems.

Literature Review

Our findings from the literature review showed that most technical standards such as ZigBee, Z-Wave, Bluetooth, ISO/IEEE 11073, and others listed in the standards section of this report are already available or currently in development and applicable in the development of AAL technologies. Different organizations are working on arranging technical standards in specific frameworks. An example is the IEEE 2413—IoT architecture, which is a unified approach to the development of IoT systems and the ISO/IEC JTC 1/SC 41—IoT and related technologies that include sensor networks and wearable technologies.

The list of privacy-related standards for IoT technologies related to AAL technology was limited when the literature review was initiated in September 2017. Only 1 year later, more than 10 ISO privacy standards relevant to AAL were found under development. This evolution in standards addresses one of the concerns identified during the study—that technology evolves very quickly, and innovation is an ongoing process. Guidelines on which technologies and structures to use can very quickly become obsolete or subject to unnecessary bureaucratic intervention.

Although the literature review has resulted in a list of existing technical standards, or under development, for use in the scope of AAL technologies, nontechnological issues have the most gaps in terms of standardization. Issues related to the human interface, processes and methods, vocabulary and social and cultural norms require special attention. Currently, there is no common terminology; hence, finding common terminology is one of the most critical areas to be consolidated in the AAL domain. Similarly, it is necessary to identify the requirements for all possible use cases, the need for specific human processes and interfaces, and create appropriate standards for each scenario. Furthermore, user engagement is widely accepted as an essential concept in the development of new systems and technologies and should be extended to the development of AAL technologies.

The design process of new AAL technologies requires special attention due to the cognitive, perceptual, or physical limitations of the target users. Such technology should not depend solely on the user’s effort and input but rather create automated solutions with minimal interaction. Yet, it is essential to take privacy considerations and concerns seriously and discuss these issues with users. Incorporating the end user in the early design stages is critical to increasing the acceptance of technology, and it is an essential part of avoiding unexpected user experience conflicts.

In this field, it is crucial to understand that the concept of end users is not limited to patients, older adults, and people with disabilities or specific health problems. Users also include therapists, health care providers, physicians, and family members who support the daily routines of vulnerable populations through the use of technology.

Interviews and Workshops

When talking to interviewees about standards and challenges in the AAL environment, the need to go beyond the technical aspect became clearer. Some challenges include creating goal-oriented and user-friendly solutions, understanding the user’s needs, and choosing the right technologies to meet those requirements. AAL technology is directly related to the intended use of the device. Each specific use case may require different details, standards, design, security, access, and data sharing. The same device that is usually considered consumer goods may, in another scenario, serve as an assistive device to an elderly or a vulnerable individual if it can improve their quality of life. Therefore, the definition of AAL technology becomes a challenge, as many consumer technologies could have AAL applications. For example, Google Home, a smart-home system, is not immediately identified as AAL technology, but it can also be used by individuals with special needs to control light switches because they cannot reach them. In this case, there is the adaptation of existing technology used to address a particular need. Adapting technology that is not designed for a specific use could put the end user at risk and compromise safety and trust.
In addition to the consideration for the end user, the purpose of the technology, accessibility, and privacy, interoperability remains to be one of the main challenges of AAL technology. The integration of products from different manufacturers through common standards will not happen without significant effort from governments and standardization agencies. Furthermore, the use of data collected at a population level for public health analysis and improvement of overall health has the potential to provide value to the data currently being collected by multiple devices. This analysis has the potential to aggregate individualized data and extracts meaning. Innovators should focus on making raw data more understandable and relevant to users and clinicians by providing context for the collected data.

It is necessary to address the concern of whether the users, health care providers, family members, and technology itself are collecting and storing data correctly, securely, and with sufficient data quality for clinical use. The creation of guidelines to ensure data reliability, trust of the data source, and trust in the process of aggregation and analysis will be critical to enable the integration of AAL technology data into clinical practice. Another critical issue is individual literacy. The focus needs to be on educating the public about AAL technology and getting them to be aware of the benefits of existing solutions in the market. Educating target users, influencers, health care providers, and the local community to guide families to better understand AAL technology and its uses can be a viable solution.

In summary, ethics, user friendliness, user acceptance, economic benefit, legal challenges, and data privacy have to be considered to provide sustainable and well-accepted AAL solutions in Canada. Additionally, more interorganizational collaborations and user-focused studies are necessary to explore the benefits of AAL technology in Canada.

**Conclusions**

Although adopting a set of standards may not address all of the gaps identified in this paper, they are essential tools that can be combined with regulations, policies, and programs to promote change. Various opportunities have been identified in this report through an extensive literature review and stakeholder consultations through interviews and workshops. User friendliness, user acceptance, and data privacy have to be considered to provide sustainable and accepted AAL solutions in Canada. Interorganizational collaborations and user-focused studies are necessary to explore the benefits of AAL technology for Canadian citizens and to ensure that this technology makes a significant positive impact on our health care system. Further in-depth research is needed to explore the existing gaps in AAL technologies.

**Acknowledgments**

The authors would like to thank all the study participants who volunteered their time to help with the development of the AAL landscape in Canada. They also acknowledge the financial support provided by the Mitacs Accelerate program. The authors would also like to acknowledge the support of members from the Ubiquitous Health Technology Lab at the University of Waterloo.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Interview questions.

[DOCX File, 15 KB - mhealth_v8i6e15923_app1.docx]

**References**


Abbreviations

AAL: active assisted living
AmI: ambient intelligence
AT: assistive technology
IEC: International Electrotechnical Commission
IEEE: Institute of Electrical and Electronics Engineers
IoT: Internet of Things
ISO: International Organization for Standardization
SyC AAL: Systems Committee on Active Assisted Living

Edited by G Eysenbach; submitted 19.08.19; peer-reviewed by H Pigot, K Gordon, M Mulvenna; comments to author 20.09.19; revised version received 27.01.20; accepted 22.02.20; published 22.06.20.

Please cite as:
URL: https://mhealth.jmir.org/2020/6/e15923
doi:10.2196/15923
PMID:32568090

©Laura X. Fadrique, Dia Rahman, Hélène Vaillancourt, Paul Boissonneault, Tania Donovska, Plinio P Morita. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 22.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Perceptions of Home Telemonitoring Use Among Patients With Chronic Obstructive Pulmonary Disease: Qualitative Study

Sara Lundell1, PhD; Mari Modig1, MSc; Åsa Holmner2,3, PhD; Karin Wadell1,4, PhD

1Division of Physiotherapy, Department of Community Medicine and Rehabilitation, Umeå University, Umeå, Sweden
2Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden
3Department of Radiation Sciences, Umeå University, Umeå, Sweden
4Division of Medicine, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden

Corresponding Author:
Sara Lundell, PhD
Division of Physiotherapy
Department of Community Medicine and Rehabilitation
Umeå University
901 87 Umeå
Sweden
Phone: 46 907868042
Email: sara.lundell@umu.se

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is a major health problem and an economic burden globally. There is growing interest in how electronic health (eHealth) can be used to provide efficient health care. Telemonitoring, where the patient’s health-related data is transmitted to a health care provider, can be used to detect early signs of exacerbations. A successful implementation of telemonitoring systems into clinical practice requires in-depth knowledge of the users’ preferences.

Objective: The aim of this study was to explore perceptions of the use of a home telemonitoring system among patients with COPD.

Methods: Semistructured individual interviews were carried out with 8 women and 5 men who were participants in a project aimed at developing and evaluating a telemonitoring system. The web-based telemonitoring system measured pulmonary function, subjective symptoms, and oxygen saturation. Participants were interviewed after having used the system for 2-4 months. Interview transcripts were analyzed with qualitative content analysis.

Results: The analysis resulted in the theme A transition toward increased control and security and four categories: using with (in)security, affecting technical concern or confidence, providing easy access to health care, and increasing control over the disease. The participants reported various perceptions of using the telemonitoring system. They expressed initial feelings of insecurity, both in terms of operating the system and in terms of their disease. However, the practical management of the telemonitoring system became easier with time; the participants gradually gained confidence and improved their self-management. New technology was perceived as an important complement to existing health care, but the importance of maintaining a human contact in real life or through the telemonitoring system was emphasized.

Conclusions: This study captured a transition among the participants from being insecure and experiencing technical concerns to acquiring technical confidence and improving disease management. Telemonitoring can be a valuable complement to health care, leading to increased self-knowledge, a sense of security, and improved self-management. Suggestions to improve the further development and implementation of telemonitoring systems include better patient education and the involvement of end users in the technical development process. Additional research is needed, particularly in the design of user-friendly systems, as well as in developing tools to predict which patients are most likely to find the equipment useful, as this may result in increased empowerment, improved quality of life, reduced costs, and a contribution to equity in health.

(JMIR Mhealth Uhealth 2020;8(6):e16343) doi:10.2196/16343

KEYWORDS
COPD; qualitative content analysis; eHealth; chronic disease; home-based care; empowerment; information and communication technology
**Introduction**

Chronic obstructive pulmonary disease (COPD) is a major public health problem [1]. According to the World Health Organization, COPD is the third-leading cause of death worldwide [2]. Although chronic progressive dyspnea is the most characteristic symptom, COPD is considered a multicomponent disease with many systemic consequences, for example, cardiovascular diseases and decreased physical capacity [1]. In addition, over 80% of patients with COPD have at least one chronic comorbidity [3].

 Patients with COPD often present with acute exacerbations (AEs) [4]. An AE is defined as an acute worsening of respiratory symptoms resulting in additional therapy [5]. Severe AEs requiring hospitalization result in a decrease in physical activity level and physical capacity with decreased muscle strength, increased impairment, and increased mortality [4,6]. Health care costs for COPD, in which costs for AEs are a strong contributing factor, are 68% higher than for people without COPD [7]. It is, therefore, of great importance to find strategies to reduce the number and severity of AEs. The annual societal costs of COPD in Sweden were estimated to be €1.5 billion in 2010 [8]. Since a history of previous AEs is an important predictor of future AEs [9], the health care target should be focused on the prevention of the first AE [1].

There is growing interest in how electronic health (eHealth) can be used to provide efficient health care. One example is telemonitoring, defined as the use of information and communication technology for the automated transmission of health-related data from a patient’s home to a health care provider. In this way, clinicians can be alerted if any abnormal parameters occur and can take immediate action to prevent complications and in-hospital care [10]. Previous studies on telemonitoring for patients with COPD have shown that telemonitoring can detect exacerbations and reduce the number of hospitalizations, as well as improve their mental-health quality of life, while the results on other health care utilization outcomes are inconsistent [11,12]. Telemonitoring seems to reduce health care costs [12,13]; however, no effect on mortality has been found [12]. A systematic review of mainly quantitative studies has shown that patients with COPD were satisfied and experienced that the telemonitoring systems were a help in monitoring their disease [14]. Reported perceptions were mixed in previous qualitative studies, ranging from experiences of the telemonitoring systems being reassuring, encouraging, and improving the self-management of the health condition to perceptions that the systems were disturbing and caused worry [14-19]. However, some of these studies have combined telemonitoring with patient education, exercise, or an action plan including medication. Consequently, few qualitative studies have explored how people with COPD experience the sole use of telemonitoring in addition to conventional care.

Some studies using telemonitoring in patients with COPD have reported a low adherence and a high dropout rate, often because of the telemonitoring system itself [14]. It is essential to understand the patients’ perspectives in order to recognize barriers to and enablers of accepting new technology for telemonitoring systems to be successfully implemented into clinical practice. Hence, there is a need for further research to ensure that telemonitoring systems fit the needs and preferences of the users. The aim of this study was, therefore, to explore perceptions of the use of a home telemonitoring system in patients with COPD.

**Methods**

**Study Design**

This study had a qualitative research design with semi-structured interviews. In order to improve transparency and strengthen transferability [20], the study was conducted and reported according to the Standards for Reporting Qualitative Research, a 21-item checklist by O’Brien et al [21].

**Setting**

This qualitative study is part of a telemonitoring project aimed at introducing and evaluating a web-based telemonitoring system that measures pulmonary function (ie, inspiratory capacity [IC] and forced expiratory volume in 1 second [FEV1]), subjective symptoms using the COPD assessment test (CAT), [1] and oxygen saturation (see Figure 1) [22,23]. The telemonitoring project was conducted in Västerbotten county in northern Sweden, a large and sparsely populated area with long distances to health care facilities for many inhabitants. The aim of the telemonitoring project was to evaluate the ability of the system to detect early signs of AEs in patients with COPD. The plan was to connect the system to health care professionals in a later stage, to detect and act on changes in symptoms indicating exacerbations. The participants used the system to perform measurements at home twice a day, 3 days a week, for 4-6 months. The system provided direct visual feedback showing heart rate, oxygen saturation, and a visual cue indicating satisfactory airflow during IC measurements. The complete spirometry curve was shown during the FEV1 procedure, but no values were calculated. No historic data were shown to the participants. Recorded data were stored locally and were automatically transmitted to the study center using the mobile 4G network. Technical support was provided to the participants during the study by one of the system developers, an electronics engineer. To evaluate if exacerbations could actually be detected, the researchers did not act on the data (ie, did not intervene with the participants during the study) unless an obvious mistake was detected. This decision was made to obtain unaffected data during the study to evaluate if exacerbations can actually be detected. The participants were well informed about this strategy and were instructed to contact their usual health care provider if they needed medical support.
**Sampling and Participants**

Participants for the interview study were recruited from a group of patients participating in the telemonitoring project [23] who were being treated for COPD at the Department of Medicine, Division of Respiratory Medicine and Allergy, University Hospital, Umeå, Sweden. The participants were approached separately for the interview study and informed that their decision would not influence their participation in the telemonitoring project or their usual care. A convenience sampling method was used, striving for a broad representation regarding age, gender, disease severity, and computer experience. Inclusion criteria for the interview study were as follows: COPD 2-4 according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification [1], aged 40 years or older, resident of Västerbotten county, included in the telemonitoring project, able to understand and speak Swedish, and have used the system for approximately 3 months. The exclusion criterion was comorbidity that could prevent participation in the interview (eg, severe psychiatric disorder). Eventually, all participants in the telemonitoring project [23] were invited and accepted to participate in the interview study.

Before enrollment in the telemonitoring project, the participants performed standard tests and received oral and written information as well as practical training in using the telemonitoring system [23]. Background data for this study was collected using the modified Medical Research Council (mMRC) dyspnea scale [24], the Hospital Anxiety and Depression Scale (HADS) [25], and a spirometer test. A question about computer experience was also included; it was answered on a scale from 1 (no experience) to 10 (very experienced). In addition, a question about how sure the participants were about their ability to use the system was answered on a scale from 1 (not sure at all) to 10 (very sure). A total of 13 participants were included: 5 men (38%) and 8 women (62%) (see Table 1).
Table 1. Characteristics of participants included in the interview study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (N=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>8 (62)</td>
</tr>
<tr>
<td>Men</td>
<td>5 (38)</td>
</tr>
<tr>
<td>Age (years), median (min-max)</td>
<td>70 (48-80)</td>
</tr>
<tr>
<td>GOLD(^a) classification grade, n (%)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8 (62)</td>
</tr>
<tr>
<td>3</td>
<td>3 (23)</td>
</tr>
<tr>
<td>4</td>
<td>2 (15)</td>
</tr>
<tr>
<td>FEV(_1)(^b)/FVC(^c) (%), median (min-max)</td>
<td>54 (25-65)</td>
</tr>
<tr>
<td>FEV(_1) (% predicted), median (min-max)</td>
<td>50 (21-68)</td>
</tr>
<tr>
<td>mMRC(^d,e) dyspnea scale (score), n (%)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3 (23)</td>
</tr>
<tr>
<td>2</td>
<td>5 (38)</td>
</tr>
<tr>
<td>3</td>
<td>1 (8)</td>
</tr>
<tr>
<td>4</td>
<td>4 (31)</td>
</tr>
<tr>
<td>HADS(^e,f) (score), median (min-max)</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>5 (0-10)</td>
</tr>
<tr>
<td>Depression</td>
<td>4 (1-12)</td>
</tr>
<tr>
<td>Computer experience(^g), median (min-max)</td>
<td>6 (1-10)</td>
</tr>
<tr>
<td>Expected ability to use the system(^h), median (min-max)</td>
<td>8 (5-10)</td>
</tr>
</tbody>
</table>

\(^a\)GOLD: Global Initiative for Chronic Obstructive Lung Disease.
\(^b\)FEV\(_1\): forced expiratory volume in 1 second.
\(^c\)FVC: forced vital capacity.
\(^d\)mMRC: modified Medical Research Council.
\(^e\)A higher number indicates increased breathlessness, anxiety, or depression.
\(^f\)HADS: Hospital Anxiety and Depression Scale.
\(^g\)Computer experience was rated on a scale from 1 (no experience) to 10 (very experienced).
\(^h\)Expected ability was rated on a scale from 1 (not sure at all) to 10 (very sure).

Data Collection
The semistructured interviews were performed between March 2014 and January 2016, after 2-4 months’ use of the telemonitoring system. The interviews were conducted by SL, typically in the participants’ homes (7/13, 54%); the rest were conducted by telephone (4/13, 31%) or at the hospital (2/13, 15%). A total of 2 out of 13 (15%) participants chose to have a family member present during the interview. The interview guide included questions about the participants’ experiences of using the system, areas of improvement, and their contact with health care providers (see Textbox 1). Each question area started with an open-ended question, then follow-up questions, and prompts were used when needed. The interviews lasted between 10 and 46 minutes, most being around 30 minutes; were audio-recorded; and were transcribed verbatim by SL or a professional transcriber.
Textbox 1. Interview guide based on question areas.

Experiences of using the system:
- What are your experiences of using the system?
- What has been good and bad?
- Has the perception of your symptoms changed?

Areas of improvement:
- What could be made better?
- How should an ideal system work?

Contact with health care:
- What health care contacts do you have?
- Have your contacts with health care providers changed?

Data Analysis
Qualitative content analysis with an inductive approach was used to analyze the data, according to the procedure described by Graneheim and Lundman [20]. The analysis was performed by MM in close collaboration with SL. The transcribed interviews were read several times to get a sense of the material, and the transcriptions were then inductively coded. The codes were compared to find similarities and differences and then grouped into categories and subcategories. This process was iterative, going back and forth, comparing categories, codes, and interview texts. Finally, an interpretative theme that could be identified across categories was abstracted from the underlying (ie, latent) content. The analysis was discussed and re-evaluated several times between MM, SL, and KW in order to achieve agreement among the researchers [20]. The three authors involved in the analysis were all physiotherapists but still contributed with different perspectives. The software Open Code 4.03 (Umeå University, Sweden) [26] was used as a tool in the analysis process to facilitate administrating the interviews, codes, and quotes.

Ethical Considerations
The study was approved by the Regional Ethical Review Board in Umeå (Dnr: 2013-187-31Ö) and followed the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants. Participant confidentiality was ensured throughout the whole process, from data collection to the presentation of the findings.

Results
A Transition Toward Increased Control and Security
The theme A transition toward increased control and security was formulated during the analysis. Participants expressed initial feelings of insecurity, both in practical aspects of using the telemonitoring system as well as regarding their disease. They experienced increased self-knowledge and deeper understanding of symptom variability when the system confirmed their health status, but it was also important to receive support from health care professionals. The practical management of the system became easier with time and participants reported an increase in technical confidence. The system was considered an important complement to existing health care, despite its lack of feedback on measurements taken during the study. However, maintaining human contact was considered essential. Overall, the use of the telemonitoring system improved the participants’ perceived self-management over time.

The theme comprises four categories and 10 subcategories (see Table 2). The categories are presented in the following section with the subcategories embedded in the text, exemplified with quotes from the participants.

Table 2. Categories and subcategories of the theme A transition toward increased control and security.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Subcategories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using with (in)security</td>
<td>Insecurity in management</td>
</tr>
<tr>
<td></td>
<td>Need for information</td>
</tr>
<tr>
<td></td>
<td>Easy to use</td>
</tr>
<tr>
<td>Affecting technical concern or confidence</td>
<td>Negative feelings about technology</td>
</tr>
<tr>
<td></td>
<td>Increased technical confidence</td>
</tr>
<tr>
<td>Providing easy access to health care</td>
<td>Complement to health care</td>
</tr>
<tr>
<td></td>
<td>Importance of human contact</td>
</tr>
<tr>
<td>Increasing control over the disease</td>
<td>Feeling reassured</td>
</tr>
<tr>
<td></td>
<td>Increased awareness</td>
</tr>
<tr>
<td></td>
<td>Management of the disease</td>
</tr>
</tbody>
</table>
Using With (In)Security

This category represents the participants’ experiences of using the telemonitoring system, reflecting both insecurity and satisfaction. Participants expressed both insecurity and the need for more support about the management of both the technology and tests, while at the same time describing the system as easy to use.

Participants experienced insecurity in management of the system and a fear of damaging it. Questions arose regarding whether the information was sent or not, expressed as follows:

But the only thing I’ve maybe felt sometimes...I wonder whether they received the information? [Woman, 58 years]

In addition, participants found it worrying not knowing if the system would work properly or not:

So that’s always at the back of your mind—I hope it works today. [Woman, 71 years]

Participants also reported that when technical failures did occur, feelings of frustration, anger, irritation, and disappointment arose. Experiencing technical failures led to insecurity and questions about whether there was an error in the system itself, or if the failure was caused by inadequate management. Furthermore, a need for simplification of the system was requested. First, the language was sometimes perceived as difficult to understand, which made the task of interpreting and answering questions challenging. Second, it was considered time-consuming and physically demanding to perform the tests, especially in the mornings; this was expressed as follows:

It is quite exerting when you try so hard, I can get chest pains afterwards. [Woman, 48 years]

This impacted daily life in a negative way. Participants reported a feeling of being “chained” to the system, since thoughts about the system and remembering to perform the tests were always present. This could lead to a feeling of being “fed up” and a wish for “a break” from the system. Since the telemonitoring system was considered “bulky,” it was perceived as difficult to take when traveling.

Participants expressed a need for information, as insufficient information about the system was raised as another area for improvement. A common problem identified was that information once provided was easy to forget. Hence, the desire for repeated information, both orally and in writing, and the desire to test the system under supervision for a longer period of time were also expressed. The need for more information justifying the use of the telemonitoring system was needed, since the purpose of using it was unclear in some cases. This was expressed as follows:

The system is good if it is useful. [Man, 73 years]

Interpreting the test results was also perceived as difficult, including both the numbers and the charts, which made it difficult to notice if there had been any changes. A desire to receive information on how normal, healthy individuals perform in these tests compared to individuals with COPD was implied:

I know I have poor expiration values, but just how poor and compared to someone who’s healthy, you can’t see that on these graphs. [Woman, 48 years]

In contrast, the system was considered easy to use, easy to learn, and time efficient, implying a general contentment among the participants. The information and practical guidance provided when the system was received were perceived as sufficient, since the equipment itself was regarded as rather self-explanatory. Furthermore, the questions about symptoms were considered comprehensible and quick to answer. Technical management did not cause any problems, and the few technical failures that occurred were interpreted as being caused by “misuse.” The opinion was that the system worked in the way that was expected, and it was difficult to suggest any areas for improvement:

It’s to say that the system itself worked perfectly, there were no problems. [Woman, 48 years]

Participants even wished to continue using it after the study was completed.

Affecting Technical Concern or Confidence

This category covers a process starting from uncertainty and negativity concerning technology and leading to increased technical confidence and interest. Participants’ initial uncertainty shifted to increased technical confidence and the ability to manage the system. New positive experiences and increased self-knowledge were expressed at the same time as negative experiences of uncertainty about new technology.

Participants expressed negative feelings about technology, describing concerns about the increasing amount of technology in society and health care. Learning new technology was perceived as difficult and new technology was even described as follows:

Terrible...I don’t know anything whatsoever about computers. [Woman, 71 years]

An interest in technology was perceived as essential in order to learn new technology. However, some expressed an unwillingness to familiarize themselves with technology and a general skepticism toward new technology, particularly in relation to health care. Furthermore, the ability to master new technology was considered to decrease with increasing age. The telemonitoring system also generated thoughts on confidentiality when it came to transmitting data electronically. Sending information electronically was considered risky and concerns were raised regarding possible cyberattacks, imposters, and information leaking, and consequently, paying bills online was, for some, unthinkable.

Others reported how using the telemonitoring system increased technical confidence. With increasing use and familiarity, the system became easier to use and worked better:

But now I’ve got a routine for it, so I think it works really well. [Woman, 58 years]

Integration of the telemonitoring system into real life seemed easier with routine. Sending data regarding health parameters electronically was considered quite harmless:
An attitude of wanting to “keep up” with technological development seemed to result in an increased interest in technology and, subsequently, a lower threshold for using it:

You have to learn because, even if you’re not interested, you just have to learn, otherwise you won’t keep up. [Woman, 48 years]

Participants reported that they had considered or had already decided to purchase a tablet.

Providing Easy Access to Health Care

A positive approach to technology marked this category. The home telemonitoring system was perceived as a promising help in existing health care, in terms of using resources properly and, subsequently, minimizing unnecessary health care visits. However, the importance of human contact was still emphasized.

The home telemonitoring system was primarily perceived as a complement to health care. In addition, participants appreciated the fact that the number of journeys, often perceived as long and stressful, could be decreased:

You can do it like this and then not have to go away [to the hospital or health care center]. [Woman, 76 years]

Distance technologies were not considered separate from ordinary care but integrated into it. Participants considered technology on the rise and visualized an opportunity to complement the telemonitoring system with video consultation, perhaps providing even better access to health care. Distance technologies were perceived to enable self-management when symptoms worsened and to provide easy communication with health care professionals. The telemonitoring system was believed to be able to help other patients with COPD in the future:

Because if you think about those who maybe live further inland and don’t have a hospital close by, then I think that it’s...They can feel more secure having a computer system like this at home. [Woman, 48 years]

The importance of human contact was emphasized, and participants expressed the need for a health care professional who can easily be contacted for counseling and guidance. There was some concern that the increased usage of technology might result in decreased human contact:

That’s terrible. Because I want someone to talk to; I want an answer when I ask about something. [Woman, 71 years]

In addition, access to a multi-professional team and being able to contact different health care professionals was desired:

That you’ve got this...that you’ve got a contact net with several people. [Woman, 58 years]

However, a face-to-face contact was not the most important, but rather the knowledge that there was a human contact on the other side:

Knowing that you can get in touch either by telephone or the internet. [Man, 72 years]

Increasing Control Over the Disease

This category captures the belief that telemonitoring could provide increased security and control over the disease. When the participants felt confident in managing the system and interpreting the results, their management of the disease also improved. They believed that receiving support from health care professionals would further contribute to increased security and control.

Participants felt reassured by the feeling of being “looked after” by health care professionals. They did not perceive it as certain that they themselves would notice a deterioration in symptoms, but it was expected that the telemonitoring system would, which made them feel secure. Feedback from health care professionals was believed to have the potential to increase their sense of security and decrease worries about data not being transmitted. To be under surveillance was said to contribute to a feeling of being reassured, safe, and comforted. They felt reassured by someone else monitoring the physical parameters and getting in touch if there were any signs of worsening:

Yes, it would almost be nice if there’s someone there...it would give a feeling of security. [Woman, 76 years]

Feedback was considered extra important at the beginning of the telemonitoring period:

I would maybe like that they sent me a graph of the results so I could see how the last month has been. Sometimes it would be nice to hear that...My god this looks really good. It works. [Woman, 58 years]

The way feedback was delivered was not considered important as long as it was provided.

Participants reported an increased awareness of the variability of symptoms and expressed that telemonitoring could confirm how they felt on better or worse days:

I can see in clear text how things are. [Woman, 61 years]

Additionally, increased confidence could result when the system confirmed the feeling of being stable in the disease:

But also when it confirms that I’m better and so on, I become more active. That I dare do more. [Woman, 58 years]

However, it was experienced that the system could also cause insecurity, for example, when the saturation device showed abnormally low values. The monitoring of symptoms was perceived as making it easier for the individual to understand how the body responds to physical tasks. Most participants found monitoring their symptoms interesting and fascinating. Measuring oxygen saturation was even performed more often than necessary out of curiosity, as this parameter was already familiar and considered interesting. The system was also identified as enabling the early treatment of AEs, which was found valuable:
So you get a little hint that something’s wrong. [Man, 72 years]

Consequently, increased awareness could impact on the management of the disease, for example, by adjusting pharmacological treatment:

I had a very high pulse during an earlier period and a low saturation level. Then I started taking cortisone and antibiotics and got better. [Woman, 58 years]

The decision whether or not to seek medical help could be facilitated by the home telemonitoring system. On the other hand, some reported no change regarding the interpretation of symptoms and activities of daily living. A recurrent trait was the reluctance to seek medical help, expressed as the following:

I’ve waited almost until I’m really bad. [Woman, 58 years]

The reluctance was explained by emotions, such as not wanting to be an inconvenience or to trouble health care professionals. Participants reported that their decision to seek medical care earlier could be strengthened when the system confirmed a deterioration in symptoms.

Discussion

Principal Findings

This study explored perceptions of the use of home telemonitoring in patients with COPD. In order to develop and implement technology that patients accept, it is important to understand barriers and enablers from the patient perspective. The main result of this study is represented by the theme A transition toward increased control and security. The theme symbolizes a process from insecurity and technical concerns to improved technical confidence and disease management, which is especially interesting considering that the participants did not get feedback from either the system or health care professionals. Thus, the study predominantly describes positive attitudes toward telemonitoring in patients with COPD and brings valuable information to the process of designing new or improved telemonitoring devices suitable for this important patient group.

Interpretation of Findings

Our findings showed, despite various previous computer experiences, a general increase in technical confidence and a newfound interest in technology with participants expressing plans for purchasing new technology such as tablets. Studies on eHealth interventions in patients with COPD and other chronic conditions have found several barriers to acceptance, such as technology anxiety, a need for technical support, insufficient information, associating the use of eHealth with dependency and ill health, and concerns that the use would influence their existing health care services [27,28]. Similar to our findings, Williams et al [16] have found that, despite initial concerns, patients were able to use a telemonitoring system effectively, regardless of previous experience. As in our study, those participants also received minimal training in using the equipment prior to onset. However, there are a few differences between our study and theirs [16]. Their interviews were conducted after 6 months, instead of 3 months as in our study, which gave the participants a longer period of time to use the telemonitoring system. Their app included a symptom diary and pulse oximeter as well as multimedia educational and self-management materials [16]. Our use of a spirometer provided additional health parameters but could require extra effort on behalf of the participants. However, the findings from our study are promising since the app was mainly considered easy to use despite the extra physical effort required by the spirometer.

Increased control of the disease was part of the transition found in our study. Several other studies have also reported that telemonitoring and other eHealth solutions can improve health knowledge and produce a sense of security for patients who know that their disease is monitored [14,15,18,27,29]. In addition, the value of being under the supervision of a health care professional has been emphasized in some of these studies [15,29]. Huniche et al [15] have also found that the monitoring of symptoms confirmed the participants’ own subjective feelings, which seemed to increase their internal resources to respond to symptoms and then facilitated the decision about whether or not to seek health care. In this study, this increased awareness and security affected the participants’ self-management, which has also been found by studies of other eHealth solutions [27,29]. Since telemonitoring has been shown to have the ability to detect exacerbations [12], improved self-knowledge and self-management are important factors for decreasing health care utilization. As in this study, Korpershoek et al [30] have stated that eHealth interventions should not replace the patients’ own feelings or undermine their decisions. On the contrary, these interventions should only confirm an underlying feeling of being okay or not, thus functioning as a support to improve self-management [30]. Our participants appreciated the possibility of detecting AEs early and acting on them. This confirms findings from a study in which participants were reassured by the idea that the system could detect impending AEs and provide objective evidence to seek health care in time [17]. However, participants in this study questioned their own feelings of health when contrasting values were shown by the telemonitoring system. This is in contrast to the findings by Huniche et al [15] where the participants, instead, questioned the accuracy of the system. It is important that the patients can trust the system and receive information to guide their actions.

In addition to the transition toward increased control and security, an interesting finding in this study is the participants’ beliefs that telemonitoring and other eHealth solutions can be a good complement to existing health care. Other studies have also reported that various patient groups perceived that eHealth solutions can increase access to health care, and that these solutions can be an improvement, an alternative, or a complement to the existing health care system [27,31]. However, the importance of human contact was emphasized by our participants. This implies that when new technology is to be implemented in health care, acceptance may be greater when the technology is presented as a complement and not as a replacement of personal care, which is in agreement with previous studies of various patient groups [30-32]. An important enabler for the effectiveness of telemonitoring in managing...
COPD is a patient-provider relationship [33]. Continuity and direct contact with a health care professional who knew the patient and gave a prompt response to symptom deterioration was crucial for our participants’ sense of security. This has also been reported earlier in studies evaluating telemonitoring interventions that included consultations with the study center [19,29]. In order to facilitate this relationship, our participants requested video consultations as a complement to the telemonitoring system. Nissen and Lindhardt [29] have reported that patients with COPD were more relaxed and focused and felt more secure during video consultations than with visits at the outpatient clinic. In contrast, a study from 2006 [34] has reported that participants expressed a sense of alienation during video consultations, as well as problems with the patient-doctor communication, in that they felt they had not been seen and that the consultation felt artificial. However, that study was published in 2006, and technological equipment is now much more common. The use of the internet, computers, and smartphones has dramatically increased [35,36], which may suggest that the general population of today is more accustomed to video calls than in 2006. A recent review by Kruse et al [33] on telemonitoring to manage COPD states that as more service options have been added to telemonitoring devices, including video consultation and phone support, a reduction in hospital admissions due to AEs has been reported.

Our participants suggested that the telemonitoring system could result in better access to health care and replace visits to a health care setting. This would minimize strenuous travel and would benefit patients living in rural areas. This is in agreement with the results of other eHealth studies suggesting that enhanced access to care is especially useful in rural areas with restricted access to health care settings [37,38]. The need to travel has been identified as a barrier for patients with COPD, due to poor mobility, lack of transport, and cost of travel [39-41]. There may also be an environmental aspect involved. Telemonitoring generates far fewer carbon emissions than traditional health care, where patients have to travel to hospitals or health care centers [42]. Therefore, making telemonitoring a natural complement to traditional health care can make an important contribution to reducing the greenhouse gas emissions fueling climate change.

In summary, in addition to the transition from insecurity to confidence and control, the participants in this study expressed both negative and positive attitudes and emotions regarding telemonitoring. A review by Kruse et al [33] has identified multiple factors as both enablers and barriers, which further illustrates this complexity. This complexity indicates that telemonitoring, and perhaps eHealth solutions in general, should be tailored to fit different patients and marketed with different strategies in order to achieve successful implementation. Another study performed in Sweden indicates that an important factor for successful implementation is that the system should meet the patients’ perceived needs and fit their self-image [43]. Therefore, including potential users at an early stage in the development process, so-called co-design [44], could be beneficial [45]. A recent study by Tistad et al [46] concludes that the involvement of user groups can strengthen the potential for a system to be adopted into everyday life and clinical practice. With this in mind, a weakness with our telemonitoring system is that the participants were not involved in its development. Perhaps negative emotions and feelings of concern and insecurity could have been reduced by using co-design.

Strengths and Limitations

We strove for trustworthiness [20] in several ways during the process of data collection and analysis and with the description of this study. First, we tried to obtain a maximum variation sample in the process of ensuring credibility [20]. However, the recruiting process was difficult, especially when it came to achieving gender balance, and fewer men were included. It was difficult to recruit men to the telemonitoring project; they had a greater tendency to decline participation than women. Therefore, toward the end, more efforts were made to include more men in order to equalize the gender balance. The reason why more men declined is unknown. This contradicts two previous studies in which there was either no significant gender difference [47] or more females than males who declined participation [48]. A majority of the interviews were 25-45 minutes long, and all of them contributed to the depth of the material. Some of the interviews were conducted by telephone for practical reasons. Several previous studies have concluded that telephone interviews are valuable for data collection and can be equated with face-to-face interviews [49-51]. Discussions about the analysis were performed repeatedly among researchers with varying competence, methodological backgrounds, and insider and outsider perspectives, which is a strength of this study. A semi-structured interview guide was used in order to increase dependability [20]. To obtain transferability [20], we have presented a thorough description of the method used and justifications of its use, and we have fulfilled the standards required for reporting qualitative research [21]. Results are presented in text, in tables, and with supportive quotes, which also enhance transferability [20].

Implications

This study’s results imply that there are many perceived enablers as well as barriers to the implementation of telemonitoring for patients with COPD. It is important not to withhold information; the participants must understand the aim of the system and be fully informed and educated about it. Appropriate training and access to prompt support is necessary to alleviate technical insecurity.

The results of this study can hopefully assist future research on telemonitoring, particularly in the design of user-friendly systems, as well as tools to predict which patients are most likely to find the equipment useful. This study has contributed to illuminating the user perspective on a telemonitoring system, which is an important perspective in the process of successful implementation. The results of this study also indicate that telemonitoring may have a potential role as a valuable complement to existing health care. It is known that patients with COPD do not receive the recommended health care [39,52]. Considering this, it would be unethical not to further explore eHealth solutions, such as telemonitoring, since it can contribute to more equity in health care. However, it is also important to consider that eHealth is not suitable for all patients. Other
solutions to complement existing health care are also needed to reduce inequity in health care.

In summary, additional research is needed regarding telemonitoring and other eHealth solutions and their implementation, bearing in mind the complexity of the enablers and barriers. The use of co-design is a promising method. It is also necessary to examine whether the use of eHealth solutions is influenced by individual factors, such as gender, ethnicity, age, and technical experience among people with COPD. These factors have been shown to influence technology use in general [53].

Conclusions

Participants in this study described various perceptions of using telemonitoring. The results implicate a transition toward increased control and security with telemonitoring, leading to increased self-knowledge, a sense of security, and improved self-management. However, concerns were raised about technical errors, insufficient information, and the ability to cope with technology. In order to further improve the development and implementation of telemonitoring systems, several actions are needed, such as improved patient education and the use of co-design where the users are involved in the development. Furthermore, telemonitoring should be viewed as a complement to existing health care, bearing in mind the importance of human contact. If correctly implemented, telemonitoring has the potential to contribute to earlier health care contacts and, thereby, the early detection of COPD AEs, improved self-management, and equity in care for patients with COPD.

Acknowledgments

The authors would like to thank the participants who took part in this study and research nurse Eva Bergmann who helped with the recruitment of participants. In addition, we would like to thank Avvai Singam (RPT) for checking the language and translating the quotes for the manuscript. This research was supported by the Swedish Research Council (Dnr: 521-2013-3503).

Authors' Contributions

SL substantially contributed to the conception and design of the study. She was the main person responsible for the collection of data and she substantially contributed to the interpretation of data and the revision of the manuscript. SL has experience from qualitative research about patients with COPD and eHealth and was responsible only for the qualitative study in the telemonitoring project. MM was responsible for the interpretation of data; she also wrote the first draft of the manuscript and contributed to its revision. MM works in primary care with a wide range of patients, including COPD patients, and contributed with a new perspective as she was not a part of the telemonitoring project. AH substantially contributed to the conception and design of the study and contributed to the revision of the manuscript. She is working with eHealth projects in a research and development unit in the Västerbotten County Council. AH was part of the research group in the telemonitoring project. KW substantially contributed to the conception and design of the study, the collection and interpretation of data, and the revision of the manuscript. She works in specialty care with this patient group and has extensive experience from research on patients with COPD, exercise, and eHealth. KW was part of the research group in the telemonitoring project. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

AE: acute exacerbation
CAT: chronic obstructive pulmonary disease assessment test
COPD: chronic obstructive pulmonary disease
eHealth: electronic health
FEV1: forced expiratory volume in 1 second
GOLD: Global Initiative for Chronic Obstructive Lung Disease
HADS: Hospital Anxiety and Depression Scale
IC: inspiratory capacity
mMRC: modified Medical Research Council

Edited by G Eysenbach; submitted 20.09.19; peer-reviewed by C Bentley, C Kruse, L Powell; comments to author 20.11.19; revised version received 20.12.19; accepted 29.03.20; published 03.06.20.

Please cite as:
Lundell S, Modig M, Holmner Å, Wadell K
Perceptions of Home Telemonitoring Use Among Patients With Chronic Obstructive Pulmonary Disease: Qualitative Study
JMIR Mhealth Uhealth 2020;8(6):e16343
URL: https://mhealth.jmir.org/2020/6/e16343
doi:10.2196/16343
PMID:32490844

©Sara Lundell, Mari Modig, Åsa Holmner, Karin Wadell. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 03.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Evaluating Online Consumer Medication Information Systems: Comparative Online Usability Study

Stefan Sigle\textsuperscript{1,2,3}, MSc; Pilar Barriga\textsuperscript{2}, MD; Francisco Javier Correa Fernández\textsuperscript{4}, MSc; Christian Juhra\textsuperscript{1}, Dr med; Steffen Härtel\textsuperscript{2}, PhD; Christian Fegeler\textsuperscript{3,5}, Dr med

\textsuperscript{1}Department of Telemedicine, University Clinic Münster, University of Münster, Münster, Germany
\textsuperscript{2}Center for Medical Informatics and Telemedicine, Faculty of Medicine, University of Chile, Santiago, Chile
\textsuperscript{3}MOLIT Institute, Heilbronn, Germany
\textsuperscript{4}National Center for Information Systems in Healthcare, Santiago, Chile
\textsuperscript{5}Institute for Medicine, Informatics and Economy, Faculty of Informatics, University of Heilbronn, Heilbronn, Germany

Corresponding Author:
Stefan Sigle, MSc
MOLIT Institute
Im Zukunftspark 10
Heilbronn, 74076
Germany
Phone: 49 7131 1334500
Email: siglestefan@gmail.com

Abstract

Background: Medication is the most common intervention in health care, and the number of online consumer information systems within the pharmaceutical sector is increasing. However, online consumer information systems can be a barrier for users, imposing information asymmetries between stakeholders.

Objective: The objective of this study was to quantify and compare the usability of an online consumer medication information system (OCMIS) against a reference implementation based on an interoperable information model for patients, physicians, and pharmacists.

Methods: Quantitative and qualitative data were acquired from patients, physicians, and pharmacists in this online usability study. We administered 3 use cases and a post hoc questionnaire per user. Quantitative usability data including effectiveness (task success), efficiency (task time), and user satisfaction (system usability scale [SUS]) was complemented by qualitative and demographic data. Users evaluated 6 existing systems and 1 reference implementation of an OCMIS.

Results: A total of 137 patients, 81 physicians, and 68 pharmacists participated in this study. Task success varied from 84\% to 92\% in patients, 66\% to 100\% in physicians, and 50\% to 91\% in pharmacists. Task completion time decreased over the course of the study for all but 2 OCMIS within the patient group. Due to an assumed nonnormal distribution of SUS scores, within-group comparison was done using the Kruskal-Wallis test. Patients showed differences in SUS scores ($P=0.02$ and task time ($P=0.03$), while physicians did not have significant differences in SUS scores ($P=0.83$ and task time ($P=0.72$). For pharmacists, a significant difference in SUS scores ($P<0.001$) and task time ($P=0.007$) was detected.

Conclusions: The vendor-neutral reference implementation based on an interoperable information model was proven to be a promising approach that was not inferior to existing solutions for patients and physicians. For pharmacists, it exceeded user satisfaction scores compared to other OCMIS. This data-driven approach based on an interoperable information model enables the development of more user-tailored features to increase usability. This fosters data democratization and empowers stakeholders within the pharmaceutical sector.

(JMIR Mhealth Uhealth 2020;8(6):e16648) doi:10.2196/16648

KEYWORDS
online consumer medication information; online usability study; sociotechnical system; information management; interoperability; implementation science
**Introduction**

Every medical decision is dependent on information, and thus, information quality is a key aspect when accessing health-related information [1-3]. Nowadays, many people worldwide (more than 60% in Europe, 80% in the United States, and 85% of the population in low- and middle-income countries) are making use of the internet to search for information about health, medication, or medical conditions; the most frequent activity is searching for medication information [4]. This substantiates the paradigm shift toward inclusive, patient-centered health care [5] and patient empowerment [6]. Despite the fact that medication provides known benefits, adequate medication use remains a challenge for patients and providers alike [7].

Online consumer medication information systems (OCMIS) try to take on these challenges by being a source for relevant medication information among patients and providers [8]. Nevertheless, these OCMIS can also create a barrier for users that have a poor ability to read, understand, and use information to make health-related decisions; this skill is referred to as health literacy [9-11]. Moreover, the quality of information contained in a given OCMIS varies [12,13] and users may be unable to differentiate between high- and low-quality information [4]. Customized needs based on users’ preferences, skills, and knowledge are often not considered by these OCMIS [14]. This creates an information asymmetry between stakeholders, which leads to poor medication adherence, causing poorer health as well as economic issues over time [15,16].

In Chile, an emerging middle-income country in Latin America [17], the number of OCMIS within the pharmaceutical sector is increasing [18]. Within the Chilean population, 58% (and almost 90% of older citizens) take at least one type of medication, of which 88% have been indicated by a medical professional [19]. Another study reported that 30% of participants indicated that they had had to suspend a treatment because of economic reasons, which can lead to long-term health problems for citizens [20]. Governmental policies promote a rational use of medications and facilitate equal access to medications and related information through OCMIS [18,21]. However, these systems have not been evaluated for their fitness for use to date.

After a feature analysis of OCMIS as part of a systematic review [22] (Multimedia Appendix 1), this follow-up study seeks to empirically investigate the usability of OCMIS through an online usability study and simultaneously considers additional factors like health literacy.

**Methods**

**Study Design**

Implementation research studies focus on real-world scenarios and identify factors that impact the uptake of research findings across multiple levels [23]. Within the context of this study, OCMIS are understood as sociotechnical systems, and the focus of this study is the human-computer interaction. This online usability study used a two-phased approach: first, there was the preceding pretest phase, which was followed by the main phase for data collection. During the pretest phase, approximately 10% of the expected participants from each group completed the study and provided feedback to researchers about the clarity and understandability of the study contents. Comments about wording obtained during the pretest were recorded as free text in digital form. Validation was performed with 2 native Spanish-speaking expert representatives from each user group and incorporated into the study after discussion. Changes to the study material were only incorporated when both experts agreed unanimously on significance and meaningfulness. Subsequently, the unmoderated main phase was conducted online, where participants acted in an in vivo setting. After the introduction video (Multimedia Appendix 2), the two-step study process was initiated: first, users completed 3 group-specific use cases with a randomly assigned OCMIS, followed by a post hoc questionnaire about the user experience during the study (Figure 1). All contents were administered in the native language. Data about the participants’ self-assessed health literacy [24,25] and OCMIS experience, as well as demographic data, were collected.

In addition, quantitative data were collected in parallel during user interaction to evaluate task success and task completion time. Data quality for the study was assured through a token system embedded in the process of accessing the study material. Pseudonymized tracking of participants without personal reference was possible, recognizing users that were not invited initially. The study was administered to participants via a URL to a self-hosted webpage where SurveyJS [26] was used for questionnaire rendering.

Participants of this study had no incentive other than to augment their knowledge about medications and OCMIS. The ethics committee at the Faculty of Medicine of the University of Chile approved this study.
A graphical view of the study procedure is shown in a Business Process Model and Notation (BPMN). After reading the introduction and consenting to participate, the participants are randomly assigned to either the case group, which uses an online consumer medication information system (online system 1...n), or the control group, which uses the reference implementation (control system). A post hoc questionnaire was performed before concluding the study. OCMIS: online consumer medication information system.

**Selection of OCMIS**

In discussion with 2 domain experts from each user group, 6 OCMIS were identified as relevant. For patients, domain experts were head organizers of patient interest groups. Physician experts were academic professionals with expertise in public health, and pharmacist experts were represented through academic professionals. After interacting with each of the platforms, experts selected relevant OCMIS based on the information needed to fulfill typical use cases. OCMIS were categorized as online pharmacies (Farmazon [27], Pharol [28]), web presence of a traditional pharmacy (Salcobrand [29]), government-driven (Ministry of Health [MINSAL] [30]; Public Health Institute of Chile [ISP] [31]), and supplier-driven (National Health Service System/La Central Nacional de Abastecimiento [CENABAST] [32]) OCMIS. OCMIS were assigned to user groups based on a decision matrix based on their features to ensure suitability.

In addition to the aforementioned OCMIS, the reference implementation TMED (medical terminology) [33], based on an interoperable information model called Chilean Pharmaceutical Terminology [34], was part of the test bench for all user groups (Figure 2).

TMED is the result of an effort to create the first vendor-neutral, standardized, and interoperable information database using Fast Healthcare Interoperability Resources (FHIR), a standard developed by Health Level 7. The information model accommodates the Chilean pharmaceutical sector, enabling users to search for and view bioequivalent generic and brand medications as well as innovator products [22,34,35]. TMED supports identification of medication type by qualities and features and provides the possibility of grouping medications by principal active substances. However, in order to evaluate its fitness for use, use cases had to be defined.
**Figure 2.** Graphical representation of the patient (green), physician (red), and pharmacist (yellow) user groups and their assigned online consumer medication information systems. Online system types include online pharmacies (Pharol, Farmazon), a traditional pharmacy (Salcobrand), medication information aggregators (MINSAL), medication information platforms (CENABAST, ISP), and a self-developed platform (TMED). CENABAST: National Health Service System of Chile; ISP: Public Health Institute of Chile; MINSAL: Ministry of Health of Chile; TMED: medication terminology.

**Use Case Definition**

Use cases are part of requirements engineering and are a narrative description of user actions and expected outcomes [36]. They allow the derivation of a feature set that must be provided by a system to the user in order to be of use. The aforementioned domain experts from each user group were prompted to write an easy-to-understand narrative text that outlines an everyday interaction with an OCMIS from their point of view, including prerequisites such as prescriptions for patients, diagnoses for physicians, and principal active substances for pharmacists (Textbox 1). Suitable exemplary medical conditions for the use cases were consented by all 6 domain experts (Textbox 2).

**Textbox 1.** Use case definition for the patient, physician, and pharmacist user groups.

- **Patient:** finding a suitable commercial product for a prescription received from a physician.
- **Physician:** finding a suitable commercial product to prescribe for a patient based on a principal active substance indicated for a diagnosis.
- **Pharmacist:** finding a suitable commercial product to restock a pharmacy, based on the need for principal active substances issued by physicians.

**Textbox 2.** Selected medical conditions used as concrete examples for the use case.

- Atypical pneumonia, which has a growing prevalence in the Chilean population [37].
- Focal epilepsy, one of the most common neuronal diseases worldwide; the majority of individuals with focal epilepsy (80%) live in low- and middle-income countries [38].
- Hypertension, one of the most common diseases; it affects more than 3.6 million in Chile and 1.3 billion worldwide [39].

**Evaluating System Usability**

Usability evaluations are critical for assuring user acceptability when designing applications [40]. Approaches from pragmatic and academic contexts are relevant when conducting usability studies [41]. International Organization for Standardization (ISO) guideline 9241-11 includes 3 dimensions for usability: effectiveness, efficiency, and user satisfaction [42]. Effectiveness is expressed as task success, efficiency is expressed as task completion time, and user satisfaction is captured in a scoring system (eg, using the SUS).

**Task Success**

The first usability dimension was measured on 3 discrete levels: complete success, partial success, and not successful. Results were aggregated dichotomously over all 3 tasks resolved by the participant by defining anything other than a complete success as not successful. Overall success was achieved if at least 2 tasks were completed successfully by the user.

**Task Completion Time**

Task completion time in seconds was measured automatically during the study for each task and user.
User Satisfaction

User satisfaction was measured using the well-established SUS, which yields a score between 0 and 100 [43]. This nonproprietary, 10-item, 5-point Likert scale tool has been extensively validated and translated into multiple languages [44]. Although it is not ideal as a standalone metric, the literature suggests combining the SUS score with task success if possible [45]. The SUS itself can be broken down into 2 principal factors: usability and learnability [46]. OCMIS were rated by each participant using the SUS as a validated measure of learnability and user satisfaction [43].

Sample Size and Internal Consistency

A sample size calculation was conducted. Literature suggests a sample size of 12-14 as sufficient to distinguish user satisfaction reliably between websites [47]. However, a sample size calculation based on a desired margin of error of 12 points in SUS score with SD of 21 and confidence level of 90%, as suggested by the literature [48], resulted in a minimum sample size of 15 participants for each platform. Internal consistency was measured using Cronbach alpha. The literature suggests acceptable values range from .70 to .90 [49,50].

Recruitment and Data Collection

Inclusion and exclusion parameters were defined prior the study. Physicians had to have completed medical school; in Chile, this includes 2 years of practical experience in the field. Pharmacists had to have at least 1 year of professional experience. Patients were only included if they had bought medication at least once in their life. Possible participants were contacted via email invitation among special interest groups (eg, for pharmacists, invitations were sent to members of the College of Pharmaceutical and Biochemical Chemists of Chile). The data collection phase lasted 3 months and was followed up by statistical data analysis.

Statistical Analysis

Initially, group-wise statistical tests were conducted, comparing platforms in terms of task time, task success, and SUS score. If results were statistically significant, an adjusted pairwise examination was performed to identify the significantly different feature. SUS score and task time were compared between OCMIS using the Kruskal-Wallis test for independent samples to compare means. Task success was evaluated using the chi-square test in combination with a standardized Z-score residual post hoc test. The Pearson chi-square test evaluated how likely it is that any observed difference between the sets arose by chance. Its null hypothesis states that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution [51]. This study evaluated the usability for OCMIS as shown in Figure 2.

Results

Baseline Statistics

Study participants included 136 patients, 80 physicians, and 67 pharmacists. The overall response rate was 283 of 4849 contacted individuals (5.8%). Table 1 provides an overview of study participant demographics. The mean ages across the different groups ranged from 31 to 38 years. Of the 136 patient participants, 87 (64%) were female, as were 36 of the 80 physician participants (45%) and 30 of the 68 pharmacists (45%). Self-assessed health literacy (where 5 indicated optimal health literacy and 1-4 indicated limited health literacy) of the study population varied between 30% and 35% for patients and pharmacists and peaked at over 50% within the physician group. Of the 67 pharmacists surveyed, 56 (83%) had used OCMIS prior to participating in this study, compared with 62 of 80 physicians (77%) and only 80 of 136 patients (58%). All participants from all groups reported that they used the internet on a daily basis, therefore data collected on internet use was not included in the overview.

To reduce the out-of-pocket spending for patients, 68 of 80 physicians (85%) reported considering the patient’s health insurance when prescribing medications, and 57 of 80 (71%) reported considering the economic situation of the patient. When asked whether generics are bioequivalent to their respective innovator medication, 33 of 80 physician participants (41%) stated that they are equal. In contrast, only 30 of 67 pharmacists surveyed (45%) agreed that innovator drugs could be replaced with generics without concern. In addition, 24 of 67 pharmacists (36%) disagreed and 12 of 67 pharmacists (18%) stated some concerns about replacing innovator drugs with generics.
Table 1. Baseline table of the participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Patients (n=136)</th>
<th>Physicians (n=80)</th>
<th>Pharmacists (n=67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>38 (11.2)</td>
<td>31 (6.2)</td>
<td>35 (9.2)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>87 (64)</td>
<td>36 (45)</td>
<td>30 (45)</td>
</tr>
<tr>
<td>Male</td>
<td>49 (36)</td>
<td>44 (55)</td>
<td>37 (55)</td>
</tr>
<tr>
<td>Health literacy, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited</td>
<td>85 (65)</td>
<td>36 (47)</td>
<td>45 (68)</td>
</tr>
<tr>
<td>Optimal</td>
<td>46 (35)</td>
<td>41 (53)</td>
<td>21 (32)</td>
</tr>
<tr>
<td>Professional experience, mean (SD)</td>
<td>N/A</td>
<td>6.57 (6.6)</td>
<td>8.86 (7.8)</td>
</tr>
<tr>
<td>Previous experience with online consumer medication information systems, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>77 (59)</td>
<td>60 (78)</td>
<td>55 (83)</td>
</tr>
<tr>
<td>No</td>
<td>54 (41)</td>
<td>17 (22)</td>
<td>11 (17)</td>
</tr>
<tr>
<td>Are generic bioequivalent medications equal to innovator medications? n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>N/A</td>
<td>33 (41)</td>
<td>24 (36)</td>
</tr>
<tr>
<td>No</td>
<td>N/A</td>
<td>41 (51)</td>
<td>30 (46)</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>6 (8)</td>
<td>12 (18)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observations per online consumer medication information system, n</th>
<th>Farmazon</th>
<th>Pharol</th>
<th>Salcobrand</th>
<th>MINHAL</th>
<th>CENABAST</th>
<th>ISP</th>
<th>TMED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32</td>
<td>30</td>
<td>44</td>
<td>15</td>
<td></td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

aThese values represent self-assessed health literacy as captured by a single-item, 5-point Likert scale where 1-4 indicated limited health literacy and 5 indicated optimal health literacy.

bProfessional experience was measured in years since graduation from university.

N/A: not applicable.

dIf Yes was indicated, the participant had used an online consumer medication information system at least once before this study.

cNot applicable.

fMINHAL: Ministry of Health of Chile.

gCENABAST: National Health Service System of Chile.

hISP: Public Health Institute of Chile.

iTMED: medication terminology.

Task Success

The second usability measure was task success (Figure 3). Patients’ task success levels were relatively consistent, independent of which OCMIS was used, ranging from 84% (Pharol) to 92% (TMED).

On the other hand, physicians’ success was heavily platform-dependent, reaching a completion rate of just 67% on MINHAL and a 100% task success rate using TMED. Pharmacists’ task success rates ranged from 50% on the CENABAST platform to 92% on the ISP platform. TMED performance was in the middle of the group, with 75% of participants successfully completing the tasks.
**Figure 3.** Binary task success rates for online consumer medication information systems: successful (light) and not successful (dark). CENABAST: National Health Service System of Chile; ISP: Public Health Institute of Chile; MINSAL: Ministry of Health of Chile; TMED: medication terminology.

**Task Completion Time**

Median task completion time in seconds for each task is shown in Figure 4. As the 3 tasks had the same structure, we hypothesized that task times would follow a downward trend; this was confirmed overall, with the exception of Farmazon and MINSAL in the patient group, where completion times increased slightly for the second and third task. In the case of TMED, initial task times are higher than with the other systems but with later tasks, the task times approach those of other OCMIS. Physicians took the least amount of time to finish the given tasks. An aggregated comparison can be found in Table 2.

**Figure 4.** Median task completion times for patients (left), physicians (center), and pharmacists (right). Times per task 1 (dark), task 2 (lighter), and task 3 (lightest) are shown with a 95% CI. CENABAST: National Health Service System of Chile; ISP: Public Health Institute of Chile; MINSAL: Ministry of Health of Chile; TMED: medication terminology.
Table 2. Overview of task success, task time, and system usability scale scores for all user groups by online consumer medication information system.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Farmazon (n_Pat = 32)</th>
<th>Pharol (n_Pat = 30)</th>
<th>Salcobrand (n_Pat = 44, n_Phy = 39)</th>
<th>MINSAL (n_Pat = 15, n_Phy = 18)</th>
<th>CENABAST (n_Pha = 20)</th>
<th>ISP (n_Pha = 28)</th>
<th>TMED (n_Pat = 15, n_Phy = 23, n_Pha = 19)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task success rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>89.6</td>
<td>84.0</td>
<td>83.8</td>
<td>84.6</td>
<td>—</td>
<td>—</td>
<td>92.3</td>
</tr>
<tr>
<td>Physician</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>100</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>50.0</td>
<td>91.7</td>
<td>77.8</td>
<td></td>
</tr>
<tr>
<td><strong>Median task time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>50.33 (27.61)</td>
<td>60.67 (50.53)</td>
<td>51.33 (74.03)</td>
<td>63.68 (61.89)</td>
<td>—</td>
<td>—</td>
<td>64.33 (32.65)</td>
</tr>
<tr>
<td>Physician</td>
<td>—</td>
<td>—</td>
<td>50.00 (236.52)</td>
<td>61.00 (478.19)</td>
<td>—</td>
<td>—</td>
<td>56.67 (179.78)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>42.33 (42.55)</td>
<td>47.67 (31.54)</td>
<td>68.00 (33.03)</td>
<td></td>
</tr>
<tr>
<td><strong>Mean SUS score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>83.83 (15.18)</td>
<td>76.38 (19.71)</td>
<td>66.73 (23.87)</td>
<td>71.33 (24.72)</td>
<td>—</td>
<td>—</td>
<td>72.67 (15.36)</td>
</tr>
<tr>
<td>Physician</td>
<td>—</td>
<td>—</td>
<td>79.66 (15.89)</td>
<td>77.06 (22.45)</td>
<td>—</td>
<td>—</td>
<td>76.85 (17.23)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>50.63 (22.24)</td>
<td>79.81 (20.68)</td>
<td>84.87 (11.62)</td>
</tr>
</tbody>
</table>

a n_Pat: number of patients.  
b n_Phy: number of physicians.  
c MINSAL: Ministry of Health of Chile.  
d CENABAST: National Health Service System of Chile.  
e n_Pha: number of pharmacists.  
f ISP: Public Health Institute of Chile.  
g TMED: medication terminology.  
h Percentage of aggregated task success rates.  
i Not applicable.  
j The median task time is in seconds.  
k SUS: system usability scale; scores can be values between 0 and 100.

**User Satisfaction**

The third dimension of usability, user satisfaction, proved to have a very high overall internal consistency, as indicated by a Cronbach alpha value of .89 for SUS scores. With one exception each in the patient and pharmacist groups, median SUS scores were above the global average of 68 (SD 12.5) for SUS scores for websites (Figure 5) [40].
Figure 5. SUS box plots: the lower (Q1) and upper (Q3) quartile, representing observations outside the 9-91 percentile range. The diagram also shows the median observation. Data falling outside the Q1-Q3 range are plotted as outliers. CENABAST: National Health Service System of Chile; ISP: Public Health Institute of Chile; MINSAL: Ministry of Health of Chile; SUS: system usability scale; TMED: medication terminology.

Group-Wise Comparison for TMED
The observed mean SUS scores for TMED ranged from 72.5 (SD 15.36) for patients to 76.85 (SD 17.23) for physicians and 84.87 (SD 11.62) for pharmacists (Table 3). SUS scores among physicians and pharmacists indicate a potential for them to be net promoters of the platform. The SUS scores were transformed into percentiles [48], adjectives, and grades [52,53] to facilitate interpretation and groupwise comparison (Table 3).

Table 3. Transformation of TMED system usability scale scores into percentile ranks, adjectives, and grades for patients, physicians, and pharmacists.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Patients</th>
<th>Physicians</th>
<th>Pharmacists</th>
</tr>
</thead>
<tbody>
<tr>
<td>System usability scale score, mean (SD)</td>
<td>72.67 (15.36)</td>
<td>76.85 (17.23)</td>
<td>84.87 (11.62)</td>
</tr>
<tr>
<td>Percentage</td>
<td>66.9</td>
<td>88.0</td>
<td>96.6</td>
</tr>
<tr>
<td>Adjective</td>
<td>Good</td>
<td>Excellent</td>
<td>Excellent</td>
</tr>
<tr>
<td>Grade (Bangor [52])</td>
<td>C</td>
<td>B</td>
<td>A+</td>
</tr>
<tr>
<td>Grade (Sauro &amp; Lewis [53])</td>
<td>B–</td>
<td>B</td>
<td></td>
</tr>
</tbody>
</table>

Statistical Evaluation
The null hypothesis was defined as not exhibiting any differences for any of the given aspects (task time, task success, SUS score), with α=.05. Due to data skewness, normality was not assumed and subsequently only nonparametrical tests were performed.

Patients
In the patient group, the differences in SUS scores (P=.02) and task time (P=.03) across OCMIS were significant, such that the null hypothesis was rejected. Pairwise SUS score comparison revealed an adjusted significant difference for Salcobrand and Farmazon (P=.008). In addition, Farmazon and Pharol differed significantly (P=.06) in pairwise completion times. However, task success did not differ significantly from expected values (P=.91).

Physicians
For the physician group, the differences in SUS scores (P=.08) and task time (P=.72) did not reach significant levels. No consecutive pairwise comparison was conducted. However, the differences in task success proved significant (P<.001) under the chi-square test. After a within-group adjustment (α=.008), MINSAL was identified as deviating significantly (P<.001).

Pharmacists
The results from the pharmacist group indicated a highly significant difference between OCMIS for SUS scores (P<.001) and task completion times (P=.007). An adjusted pairwise comparison for SUS scores revealed a significant difference between CENABAST (P<.001) and ISP as well as CENABAST and TMED (P<.001). When focusing on completion time, only CENABAST and TMED showed significant differences (P<.005). The differences in task success among pharmacists was significant (P=.008); after a post hoc adjustment (α=.008), the CENABAST OCMIS was found to deviate from expected values (P=.004).

Qualitative Data
In addition to quantitative data, 76 of 136 patients (55%), 36 of 80 physicians (45%), and 31 of 67 pharmacists (46%) provided qualitative feedback about features that they considered desirable for OCMIS. Comments were analyzed for their content and tagged by keyword (Textbox 3).
Textbox 3. User comments on critical features for online consumer medication information systems, ranked by overall occurrence.

- The up-to-date or approximated medication price should be displayed (132 mentions).
- Search flexibility should be increased (eg, searching for principal active substances or quality parameters; 11 mentions).
- Disambiguation of search terms (eg, phonetic searches) should be provided (10 mentions).
- Medication concentrations should be displayed (6 mentions).
- Adverse effect information should be provided (6 mentions).
- An increased amount of information about medications (eg, kinetics and posology) should be included (4 mentions).
- Evidence for medications should be shown (3 mentions).
- Filters for information such as dosage or concentration should be implemented (3 mentions).
- Integration to other knowledge databases should be considered (3 mentions).
- Georeferenced information for pharmacies and stock considerations should be included (2 mentions).
- Personal discounts due to insurance coverage should be included in the price calculation (1 mention).
- Information neutrality should be a priority (1 mention).
- Native mobile applications should be preferred (1 mention).

Discussion

Principal Findings

An online usability study was conducted to evaluate OCMIS on the dimensions of task success (completion), task completion time, and user satisfaction.

The ongoing controversy of whether to prescribe innovator medications or use bioequivalent generic products is reflected within the study population. Generally, physicians are slightly more confident in using generic products than pharmacists.

For patients, online pharmacies (Farmazon and Pharol) seemed to be the most suited to their tasks as indicated by high user satisfaction scores. Task time was significantly lower for the OCMIS of traditional pharmacies when compared to online pharmacies. Task success rates indicated that all platforms seemed to be suited for the use case.

Physicians seemed to have difficulties completing their tasks when using the MINSAL platform, but not when using the OCMIS of traditional pharmacies (Salcobrand) or the reference implementation (TMED).

The user satisfaction scores of pharmacists identified both ISP and TMED as the most usable platforms, with no significant difference in user satisfaction between them. The platform of public medication supplier CENABAST received lower SUS scores and also had lower task success rates.

Strengths and Limitations

For the selection of OCMIS, a discussion with 2 professional representatives was conducted; this may not be representative of which OCMIS are used by health care professionals on a national level. However, more than half of the participants indicated an awareness of the OCMIS presented in this study, indicating that the selected OCMIS were relevant. Health literacy was not homogeneous among participants, indicating unequal starting conditions for each participant; however, this reflects reality. Participant recruitment was carried out by email distribution to special interest groups, which might introduce bias as these individuals may have a higher awareness of OCMIS.

Due to the design of online usability studies, a unique combination of advantages was achieved. The study was not moderated and no social desirability response bias [54] was introduced by this in vivo setting, assuring the most natural conditions for the user while they evaluated the OCMIS. The study design facilitated the automated collection of qualitative and quantitative data directly after the experience. In comparison to traditional usability studies, a higher number of participants was recruited in a shorter time frame, which contributed to the robustness of the results.

Conclusions

This study demonstrated that TMED is a promising approach and showed that interoperable, neutral information models can empower stakeholders in context-agnostic medication decisions. Although an independent group should verify these results to avoid any potential bias, TMED was statistically proven to not be inferior to other OCMIS in usability aspects, while offering flexible search and extension capabilities due to its underlying interoperable information model.

Based on the results and qualitative feedback on desired features provided by participants, improvements can be incorporated to alleviate information asymmetries and foster data democratization within the pharmaceutical sector even further by providing user-tailored information. The approach of personalized drug information provision is promising and can serve as a basis for other applications, such as electronic prescriptions, and enable research opportunities through its standardized approach.
Acknowledgments

The authors thank the following institutions for their support: Centro Nacional en Sistemas de Información en Salud (CORFO 16CTTS-66390), Biomedical Neuroscience Institute (ICM, P09-015-F), DAAD PAGEL CHIP: Chilean Health Info and Process Challenge (DAAD 57220037), DAAD Scholarship: binationally supervised doctorates (DAAD 57314603) and the MOLIT Institute.

Conflicts of Interest

We are responsible for the development of TMED, one of the tested online consumer medication information systems. No other conflicts of interest are declared.

Multimedia Appendix 1

Benchmarking of online consumer medication information systems. This benchmark was used in a previous publication [22].

Multimedia Appendix 2

Introduction and overview video for the online usability study.

References


47. Tullis T, Stetson J. A Comparison of Questionnaires for Assessing Website Usability. 2004 Presented at: Usability Professionals Association (UPA); June 7-11; Minneapolis, Minnesota.


51. Pearson K. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2009 Apr 21;50(302):157-175. [doi: 10.1080/1478644009463897]


Abbreviations

CENABAST: National Health Service System/La Central Nacional de Abastecimiento
FHIR: Fast Healthcare Interoperability Resources
ISP: Public Health Institute of Chile
MINSAL: Ministry of Health
OCMIS: online consumer medication information system(s)
SUS: system usability scale
TMED: medication terminology (Spanish: Terminologia de Medicamentos)
Factors Influencing Acceptance of Personal Health Record Apps for Workplace Health Promotion: Cross-Sectional Questionnaire Study

Hyun Sang Park¹,², MS; Kwang Il Kim³, MS; Jae Young Soh¹, MS; Young Ho Hyun¹; Sae Kyun Jang⁴, PhD; Sol Lee⁵; Ga Young Hwang⁴, MS; Hwa Sun Kim⁵, PhD

¹Digital Healthcare Department, BIT Computer Co Ltd, Seoul, Republic of Korea
²Department of Medical Informatics, Kyungpook National University, Daegu, Republic of Korea
³Finance Programs Department, Korea Occupational Safety & Health Agency, Ulsan, Republic of Korea
⁴Research Institute, HealthConnect Co Ltd, Seoul, Republic of Korea
⁵Elecmarvels Co Ltd, Daegu, Republic of Korea

Corresponding Author:
Hyun Sang Park, MS
Digital Healthcare Department
BIT Computer Co Ltd
BIT Building 33, Seocho-daero 74-gil, Seocho-gu
Seoul
Republic of Korea
Phone: 82 2 3486 1234 ext 507
Fax: 82 2 3486 1983
Email: hspark@bit.kr

Abstract

Background: Health care technologies can help improve workers’ health and productivity by supporting workplace health promotion. A personal health record app is used to manage medical data such as results from medical checkups, which facilitates decision making for medical personnel. However, an analysis of users’ technology acceptance is required to provide appropriate services based on personal health record apps.

Objective: The purpose of this study was to analyze the factors influencing the behavioral intention of health experts and workers to use an app in workers’ health centers and to examine differences in their perception of the main variables.

Methods: The study involved health experts and workers who visited 21 workers’ health centers in Korea to verify a research model in which perceived risk was added to the unified theory of acceptance and use of technology, a representative theory of information technology acceptance. After receiving ethical approval from the Korea National Institute for Bioethics Policy, 1050 questionnaires were distributed over 7 weeks with cooperation of the Korea Occupational Safety and Health Agency. A multiple linear regression analysis and multigroup path analysis were performed to verify the hypotheses, and independent samples t tests were performed to analyze differences between workers’ and health experts’ perception of the main variables.

Results: The analysis included data from 866 respondents (687 workers and 179 health experts). Effort expectancy (β=.08, P=.03), social influence (β=.43, P<.001), performance expectancy (β=.07, P=.008), and facilitating conditions (β=.13, P<.001) exerted significant positive effects on behavioral intention, whereas perceived risk (β=-.29, P<.001) exerted a significant negative effect on behavioral intention. Performance expectancy had a significant effect on path differences depending on gender (critical ratio=-3.38) and age (critical ratio=1.97). Workers’ mean scores for the main variables were higher relative to those of health experts for all remaining variables except perceived risk, and significant differences were observed for all remaining variables except facilitating condition.

Conclusions: Social influence exerted the strongest effect on behavioral intention to use the personal health record app. Consequently, it is necessary to coordinate health promotion activities in the workplace as well as the operational direction of community institutions such as in workers’ health centers to allow workers to manage their own health via continuous use of the app. In addition, the app should be developed based on a requirement analysis of the balance between both interest groups in consideration of differences in perspective between consumers and service providers.
Introduction

Background

Many workers spend most of their waking hours in the workplace [1], which is an environment that can have both positive and negative effects on health [2]. As such, the workplace is the best environment to apply the concept of health promotion. The World Health Organization declared that workplaces should be a priority for health promotion [3]. Workplace health promotion entails employers, workers, and communities working together to improve workers’ mental and physical health and overall well-being [4]. Elaborately designed workplace health promotion not only improves workers’ health [2] but also positively affects their productivity [5]. The primary challenge in workplace health promotion involves how to increase worker participation, given that less than 50% of participants typically remain in workplace health promotion programs [6] and the median attrition rate is 28% [7]. These obstacles can be overcome by incorporating health care technologies (eg, electronic health, mobile health [mHealth], wearable devices) into workplace health promotion strategies [8].

Health care technologies can increase workers’ interest, motivation, and participation in workplace health promotion [9,10]. These technologies function as cost-effective health promotion and disease prevention mechanisms by allowing workers to monitor their own health. Workers are at increased risk from stress caused by heavy workloads and unhealthy lifestyles, including lack of exercise and frequent drinking, relative to the general public [11]. In particular, office workers sit for long periods of time in the workplace, which exposes them to an increased risk of developing chronic diseases such as heart disease, cerebrovascular disease, and hypertension [12] that are all associated with high mortality rates as the second, third, and ninth most prevalent causes of death in Korea, respectively [13]. The cost associated with the treatment of cardiovascular disease has reached US $6.9 billion [14], which is higher than the US $4.7 billion spent on cancer, as the most prevalent cause of death, and the burden of disease in Korea is high. Moreover, mortality from cardiovascular disease has steadily increased over the past 10 years [15]. Effective prevention and management are essential because cardiovascular disease not only harms workers’ health but also increases medical expenses [16] and contributes to the social burden caused by decreased corporate productivity [17].

Workplace health promotion using health care technology has been shown to improve participants’ physical activity and eating habits [18,19]. Setting goals, supporting self-monitoring, and providing feedback on changes in physical activity and eating habits can be an effective mechanism for workplace health promotion. Various health care technologies have been studied for efficient workplace health promotion application. Mattila et al [20] conducted a 1-year randomized controlled trial to investigate the activity and usefulness of personal health technologies (web services, mobile apps, and personal monitoring devices) that support workplace health promotion. The authors showed that less than 30% of subjects continued to use mobile apps and web technologies, and that the key requirements for personal health technologies were simplicity, integration with everyday life, and clear feedback. Cook et al [9] conducted a randomized controlled trial for 3 months to evaluate the effectiveness of a web-based workplace health promotion program and found that web-based programs were more effective than a print-based intervention for improving diet and nutrition, but not for improving stress and physical activity. Balk-Møller et al [21] conducted a randomized controlled trial for 38 weeks to investigate the motivations of workers involved in web and app-based workplace health promotion and reported that social functions were more popular than personal functions, and social factors motivated continued use.

Choia et al [22] studied workers’ intention to use health care technology through an investigation of construction workers’ acceptance of wearable devices (smart vests and wristbands) for occupational safety and health based on the technology acceptance model (TAM) [23]. They found that perceived usefulness, social influence, and perceived privacy risk were related to the intention to adopt wearable devices. Mohadis et al [24] investigated office workers’ acceptance of mHealth apps designed to increase physical activity based on the unified theory of acceptance and use of technology (UTAUT) [25], and found that performance expectancy and social influence had a significant effect on behavioral intention, but not on effort expectation. Sari et al [26] proposed a UTAUT-based conceptual framework to identify factors that influence worker adoption of and intention to use mHealth technology. Technology acceptance of mHealth apps has also been tested in patients with chronic diseases [27], younger adults [28], and health care professionals [29], but related research on workers in workplaces is still in its infancy [30]. Although many mHealth apps have been developed to date, few of these apps have been developed specifically to improve workplace health promotion [31].

Interest in personal health management is rising as aging and the incidence of chronic disease increase [32]. Moreover, active services focused on prevention and health promotion are needed [33] to record lifestyle factors such as exercise, nutrition, and sleep via various wearable devices [34-36] and to measure blood pressure, blood sugar, and weight via personal health devices [37]. Occupational factors such as the workplace environment should also be considered in managing chronic diseases, either by integrating occupational information into the electronic health record (EHR) [38] or implementing the occupational data for health model [39]. Recently, Health Level Seven (HL7) designed a fast health care interoperability resource (FHIR) profile [40] to represent patients’ occupational elements in personal health records (PHRs). PHRs are electronic tools that allow secure
access, management, and sharing of health information [41], which is generally monitored by patients [42]. Individuals can check medical records provided by hospitals, monitor information regarding prescribed medicines and test results, and manage exercise and diet information related to health promotion [43]. Patients can use PHRs to reduce additional medical expenses, and disease management, treatment, and prevention activities can be enhanced as cooperation is improved through communication among medical personnel [44,45].

Employers expect workers to participate in workplace health promotion and enjoy the benefits provided by the organization. In this situation, the concept of PHRs is prominent owing to program technology-based attributes [46]. Employers can provide PHRs that motivate workers’ health care [47]. Employer-sponsored PHRs are driven by commercial goals to reduce productivity loss and health insurance costs by promoting a healthy lifestyle [48]. Despite the interest in and expected effects of PHRs, it is difficult to successfully provide these services [49]. Google Health, released in 2008, suspended the service in 2011 because of poor user participation [50], and Microsoft announced in November 2019 that they would stop providing their PHR HealthVault. However, Apple’s HealthRecord offers large-scale services that could be linked to over 200 medical institutions as of February 2019. These services are offered free of charge, but users’ technology acceptance for PHRs remains low and must be addressed [51]. To implement and operate a successful PHR app service in the workplace, employers, policymakers, developers, and planners must be aware of factors affecting workers’ technology acceptance.

Therefore, the purpose of this study was to analyze the factors that influence acceptance of PHR apps in the workplace and examine differences in perceptions surrounding the main factors. We applied a research model that included Bauer’s perceived risk [52] as an independent variable to the UTAUT [25], which is widely used to explain acceptance of new information technology.

Theoretical Background and Related Works

Unified Theory of Technology Acceptance Use

Previous studies examining the acceptance of new information technology adapted the TAM [23] to research technology. However, the TAM does not adequately support the validity of the relationships between exogenous variables, and is therefore suitable for simple technology acceptance studies but has limited ability to analyze interrelationships in complex environments. Venkatesh et al [25] proposed the UTAUT, which assesses users’ technology acceptance from an integrated perspective based on eight representative related theories, including the TAM. The UTAUT model consists of four independent variables (performance expectancy, effort expectancy, social influence, and facilitating conditions) and four moderating variables (gender, age, experience, and voluntariness of use; see Figure 1).

Performance expectancy, effort expectancy, and social influence affect behavioral intention, and facilitating conditions affect use behavior. Performance expectancy is the degree to which system use is perceived to improve work performance; effort expectancy reflects the usability of a system; social influence refers to the degree of awareness that others deemed to be important believe that one should use a new system; and facilitating conditions reflects the degree to which individuals believe that the necessary organization and technical infrastructure are in place to support the use of new systems. These variables are in turn influenced by gender, age, experience, and voluntariness of use when they affect users’ behavioral intention and use behavior.

In a previous study, the UTAUT model explained 70% of behavioral intention and use behavior for an information system, representing a significant improvement in the explanatory power of the model relative to that of existing models, which described 40% of the technology acceptance [25]. Therefore, the UTAUT model can be used to explain users’ technology acceptance of newly developed information technology in medical informatics, which actively converges with other industries. Most previous studies have included electronic medical records [53], PHRs [51], health care devices [54,55], mobile and electronic health services [24,56-59], and telemedicine services [60-63] in the UTAUT model.
**Perceived Risk**

Perceived risk, first introduced by Bauer [52], is the risk subjectively perceived by consumers when performing certain actions such as the uncertainty consumers feel when they cannot predict the outcome of purchase decisions. Short [64] demonstrated that individuals experienced the consequences of this danger; Rayner and Cantor [65] showed the probability of an adverse event occurring and examined subjective assessment of the magnitude of damage incurred by the event. Previous studies [66,67] have examined the effects of perceived risk on users’ acceptance. In addition, previous medical informatics studies [68-70] have examined users’ perceived risk of mHealth technology and electronic medical records.

**Workers’ Health Center**

According to a 2017 analysis of industrial accidents [71] conducted by the Ministry of Employment and Labor, of the 993 deaths due to occupational diseases, 354 involved cardiovascular disease, 215 of which occurred at workplaces with fewer than 50 employees. The risk of cardiovascular disease in workers could be decreased via continuous health management according to the results of medical checkups. However, workers in vulnerable classes often do not benefit from systematic industrial health services because they are rarely offered in small-scale workplaces [72,73]. The workers’ health center was established in Korea in 2011 to meet the rising need for disease prevention and health promotion services in small-scale workplaces.

Workers’ health centers are set up in areas with many small-scale workplaces such as industrial parks, and provide services for preventing occupational disease in workers. There are currently 21 workers’ health centers in operation in Korea, staffed by professional personnel such as occupational and environmental medicine specialists, occupational nurses, industrial hygiene safety engineers, physical therapists, and counseling psychologists, who provide comprehensive occupational health services, including occupational disease prevention, cerebrovascular disease prevention, musculoskeletal disease prevention, workplace environment counseling, job stress prevention, and lifestyle improvement. Workers’ health centers are used by 180,000 workers each year, most of whom are interested in their health care. Workers’ health centers manage a vast amount of worker data through the electronic worker health management system.

**PHR App**

Our PHR app (Figure 2) manages a worker’s PHRs (eg, life logs, health information, and medical checkup data) and supports customized health care services and workplace health promotion through links between specific systems and platforms. The worker PHR complies with HL7 FHIR Release 4. The app’s primary functions include data collection, text-based health counseling, consultation reservations, sharing PHRs, and viewing occupational health content. For example, workers can collect their own data stored in the workplace or at the workers’ health center through the PHR app, manage their PHRs, and receive health counseling services from health experts. In addition, by sharing PHRs to specific platforms through self-certification and consent, analysis results (eg, disease prediction, health, and body age) can be confirmed. PHRs based on data collected at workplaces are the basis for continuous health care, regardless of the worker’s external environment (new workplace turnover, local agencies, and hospitals). These data can help decision making for medical personnel by collecting and sharing data among various institutions through our PHR app, which ensures interoperability.
Methods

Research Model

The PHR app is an important element of the next generation of health care services and supports personal health promotion by storing and managing important personal medical data in one location. Moreover, using the UTAUT model to analyze and predict users’ technology acceptance of the PHR app is a rational approach. In this study, we set the main variables of effort expectancy, social influence, performance expectancy, and facilitating conditions as factors affecting behavioral intention to use the PHR app based on the UTAUT model. In the model, the dependent variable, use behavior, is affected by behavioral intention and facilitating conditions, and behavioral intention is determined by performance expectancy, effort expectancy, and facilitating conditions. However, as PHRs are currently in the introduction stage in Korea, research on actual users is limited. Therefore, in this study, we assumed that facilitating conditions also affected behavioral intention, and therefore included behavioral intention as a dependent variable without considering use behavior. Gender and age were also assumed to moderate the effects of performance expectancy, effort expectancy, and social influence on behavioral intention.

The PHR app is accompanied by various risks, which exert direct effects on behavioral intention. For example, medical data collected and utilized by the PHR app contain highly sensitive information, and behavioral intention decreases when there is a high probability of a data breach or fraud. In addition, behavioral intention decreases when users cannot securely manage as much information as they expected, or believe their information will be used for other purposes. These risks should be minimized when introducing health care services based on the PHR. Therefore, this study extended the existing UTAUT model by including perceived risk as a main variable (Figure 3).
Research Hypotheses

Performance expectancy is defined as “the degree to which an individual believes that using the system will help improve his or her performance” [25]. Performance expectancy is similar to the perceived usefulness of the TAM [23]. In this study, performance expectancy refers to the degree to which users believe that using the PHR app will help them improve their health. The PHR app not only collects and manages the worker’s own data but also improves motivation and participation for workplace health promotion through personalized feedback from health experts [74-77]. Previous studies have demonstrated that performance expectancy and usefulness affect behavioral intention [51,63,78-80]. Therefore, we proposed the following hypothesis, H1: Performance expectancy in the PHR app will exert a positive effect on behavioral intention.

Effort expectancy is defined as “the degree of ease associated with the use of the system” [25], and affects the intent to use and the ease with which users can learn and use the system, similar to the TAM’s perceived ease of use [23] and innovation diffusion theory [81]. In this study, effort expectancy refers to the PHR app’s ease of use; the app should be easy to use, taking workers’ diversity into account (eg, age, type of business). Given that hard-to-use functions have a negative impact on users [82], PHR app functions should make it easy for users to access their data. Previous studies have shown that ease of use affects behavioral intention [51,83,84]. Therefore, we proposed the following hypothesis, H2: Effort expectancy in the PHR app will exert a positive effect on behavioral intention.

Social influence is defined as “the degree to which an individual perceives that important others believe that he or she should use the new system” [25]. In this study, social influence refers to the degree to which users feel that others deemed to be important or work colleagues believe that the user should use the PHR app for enhanced health management. Social influence is associated with organizational culture in the workplace [85], which is known to be an important social characteristic that affects organization, group, and individual behavior [86-89]. Social influence is an important factor in workers’ communications and interactions with their colleagues [90], which has been identified as a suitable factor for technology acceptance in previous studies [55,91-93]. Therefore, we proposed the following hypothesis, H3: Social influence in the PHR app will exert a positive effect on behavioral intention.

Facilitating conditions are defined as “the degree to which an individual believes that an organizational and technical infrastructure exists to support the use of a given system” [25]. In this study, facilitating conditions refers to a user’s belief that an organizational and technical infrastructure exists to support PHR app use. The number of aging individuals is rapidly increasing. Older workers are less capable of acquiring and adopting new technologies such as information and communication technology, and need adequate training and technical support for using such tools. Previous studies have shown that facilitating conditions affect workers’ behavioral intention to use new technology [94-96]. Therefore, we proposed the following hypothesis, H4: Facilitating conditions in the PHR app will exert a positive effect on behavioral intention.

Perceived risk is defined as “the uncertainty due to unforeseen consequences” [52]. In this study, perceived risk refers to the degree to which users are aware of possible loss associated with uncertainty surrounding PHR app use. Since workplaces require insight into the health status of their organizations and workers, personal health data should be strictly protected [97]. The health manager in the workplace handles sensitive medical data such as workers’ medical checkups and data from individuals who have abnormal findings; the PHR app collects and manages these data. Confidence in the system, information privacy, and
Security concerns affect the sharing behavior of PHRs [98]. Therefore, we propose the following hypothesis, H5: Perceived risk in the PHR app will exert a negative effect on behavioral intention.

In UTAUT, performance expectancy, effort expectancy, and social influence are moderated by gender, age, experience, and the voluntary nature of use [25]. Gender and age are generally perceived as significant factors in attitudes toward information technology [99], which previous studies have treated as moderating variables in technology acceptance [100-102]. Therefore, we proposed the following hypotheses: H6, Effects of performance expectancy on behavioral intention of the PHR app will be moderated by gender and age; and H7, Effects of effort expectancy on behavioral intention of the PHR app will be moderated by gender and age; and H8, Effects of social influence on behavioral intention of the PHR app will be moderated by gender and age.

**Instrument Development**

The questionnaire consisted of 42 items. Responses to the 21 items (see Multimedia Appendix 1) measuring the main variables were on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree” (Table 1). In addition, 12 items pertained to participants’ general characteristics and 9 items concerned the experience of and functions required of the app.

**Table 1.** Definition of variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Number of questions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral Intention</td>
<td>The degree of users’ behavioral intention to use the PHR app</td>
<td>3</td>
<td>[25]</td>
</tr>
<tr>
<td>Performance Expectancy</td>
<td>The degree to which users believe that using the PHR app will help them to improve their health</td>
<td>3</td>
<td>[24,25]</td>
</tr>
<tr>
<td>Effort Expectancy</td>
<td>Ease of use for the PHR app</td>
<td>4</td>
<td>[24,25]</td>
</tr>
<tr>
<td>Social Influence</td>
<td>The degree to which users feel that important others or work colleagues believe that the PHR app should be used for enhanced health management</td>
<td>4</td>
<td>[22,24,25]</td>
</tr>
<tr>
<td>Facilitating Conditions</td>
<td>The degree to which users believe that an organizational and technical infrastructure exists to support use of the PHR app</td>
<td>4</td>
<td>[2,25]</td>
</tr>
<tr>
<td>Perceived Risk</td>
<td>The degree to which users are aware of possible loss relating to uncertainty surrounding use of the PHR app</td>
<td>3</td>
<td>[52]</td>
</tr>
</tbody>
</table>

**Data Collection and Analysis**

After receiving ethical approval (IRB No. P01-201902-23-014) from the Korea National Institute for Bioethics Policy, we recruited workers and health experts who visited 21 worker health centers with cooperation of the Korea Occupational Safety and Health Agency. The survey included 40 workers and 10 health experts from each worker health center and considered regional distribution.

We used a paper-based questionnaire to verify the proposed research model. The survey was conducted for approximately 7 weeks from February 12 to April 6, 2019. In total, 1050 questionnaires were distributed and 900 were collected. Of the collected questionnaires, we removed 34 that were inappropriate for analysis because they did not meet the purpose of the study or included insincere responses.

Frequency statistics were used to analyze the participants’ general characteristics. Reliability analysis, exploratory factor analysis, and confirmatory factor analysis were used to assess the instrument’s reliability and validity. Correlation analysis was used to examine bivariate associations among the main variables. Multiple linear regression analysis was employed to verify the explanatory power and hypotheses of the research model. Multigroup path analysis was performed to verify the effect of moderating variables, and critical ratios for differences were calculated to verify the statistical significance of the path coefficient for each group. Independent samples t tests were applied to analyze differences in perception of the main variables. For frequency analysis and independent samples t tests, the participants were divided into workers and health experts, and the remaining analyses were based on all respondents.

We used the factor extraction method, which implements the most commonly used maximum-likelihood estimation, in common factor analysis; for factor rotation, we used the direct Oblimin method in the square rotation to examine correlations between the factors. The determination criterion for the number of factors extracted was an eigenvalue >1. The Kaiser Meyer Olkin index was used to test the suitability of factors for analysis, where values of 0.5-1.0 indicate that the factors are suitable for analysis and those ≤0.5 indicate that the factors are not suitable for analysis. Pearson correlation analysis was performed to examine bivariate correlations before performing multiple linear regression analysis. Correlation analysis was performed using factor score storage values in the factor analysis. In the social sciences, coefficients below ±0.2 are considered low, those between ±0.3 and ±0.7 are considered moderate, and those above ±0.7 are considered high [103]. Analyses were performed with IBM SPSS Statistics ver. 25.0 and AMOS 22 (IBM Corp, Armonk, NY, USA).

**Results**

**Participant Characteristics**

The study involved 866 participants, including 179 health experts and 687 workers who visited 21 workers’ health centers
nationwide. Workers’ general characteristics are summarized in Table 2. The majority of the workers were women. Most participants were older than 50 years, followed by those in their 40s and 30s. The duration of employment in the workplace was most commonly 1-4 years, and 66.1% (454/687) of workers were employed in workplaces with <50 employees. Clerical and service-based businesses were more common than production and technical businesses.

Health experts’ characteristics are shown in Table 3. Similar to the workers, the majority of respondents were women. The most common age group was 30-39 years and the duration of employment in the workplace was most commonly 1-4 years. The most common type of occupation was nursing, followed by physical therapy, industrial hygiene safety engineering, counseling psychology, occupational and environmental medicine, and other.
Table 2. Characteristics of the workers (N=687).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>266 (38.7)</td>
</tr>
<tr>
<td>Female</td>
<td>421 (61.3)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>20-29</td>
<td>124 (18.0)</td>
</tr>
<tr>
<td>30-39</td>
<td>159 (23.1)</td>
</tr>
<tr>
<td>40-49</td>
<td>165 (24.0)</td>
</tr>
<tr>
<td>≥50</td>
<td>237 (34.5)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>207 (30.1)</td>
</tr>
<tr>
<td>Married</td>
<td>462 (67.2)</td>
</tr>
<tr>
<td>Widowed</td>
<td>10 (1.5)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>8 (1.2)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Middle school</td>
<td>34 (4.9)</td>
</tr>
<tr>
<td>High school</td>
<td>190 (27.7)</td>
</tr>
<tr>
<td>College (2 years)</td>
<td>100 (14.6)</td>
</tr>
<tr>
<td>College (4 years)</td>
<td>313 (45.6)</td>
</tr>
<tr>
<td>Graduate school</td>
<td>50 (7.3)</td>
</tr>
<tr>
<td><strong>Time in the workplace (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>111 (16.2)</td>
</tr>
<tr>
<td>1-4</td>
<td>251 (36.5)</td>
</tr>
<tr>
<td>5-9</td>
<td>138 (20.1)</td>
</tr>
<tr>
<td>≥10</td>
<td>187 (27.2)</td>
</tr>
<tr>
<td><strong>Number of employees in the workplace</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>82 (11.9)</td>
</tr>
<tr>
<td>5-9</td>
<td>105 (15.3)</td>
</tr>
<tr>
<td>10-29</td>
<td>173 (25.2)</td>
</tr>
<tr>
<td>30-49</td>
<td>94 (13.7)</td>
</tr>
<tr>
<td>50-99</td>
<td>53 (7.7)</td>
</tr>
<tr>
<td>≥100</td>
<td>180 (26.2)</td>
</tr>
<tr>
<td><strong>Type of business</strong></td>
<td></td>
</tr>
<tr>
<td>Production</td>
<td>67 (9.8)</td>
</tr>
<tr>
<td>Clerical</td>
<td>271 (39.4)</td>
</tr>
<tr>
<td>Service-based</td>
<td>226 (32.9)</td>
</tr>
<tr>
<td>Technical</td>
<td>54 (7.9)</td>
</tr>
<tr>
<td>Other</td>
<td>69 (10.0)</td>
</tr>
</tbody>
</table>
Table 3. Characteristics of health experts (N=179).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>45 (25.1)</td>
</tr>
<tr>
<td>Female</td>
<td>134 (74.9)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
</tr>
<tr>
<td>20-29</td>
<td>27 (15.1)</td>
</tr>
<tr>
<td>30-39</td>
<td>104 (58.1)</td>
</tr>
<tr>
<td>40-49</td>
<td>35 (19.6)</td>
</tr>
<tr>
<td>≥50</td>
<td>13 (7.3)</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>75 (41.9)</td>
</tr>
<tr>
<td>Married</td>
<td>103 (57.5)</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>College (2 years)</td>
<td>28 (15.6)</td>
</tr>
<tr>
<td>College (4 years)</td>
<td>95 (53.1)</td>
</tr>
<tr>
<td>Graduate school</td>
<td>56 (31.3)</td>
</tr>
<tr>
<td>Time in the workplace (years)</td>
<td></td>
</tr>
<tr>
<td>&lt;1</td>
<td>30 (16.8)</td>
</tr>
<tr>
<td>1-4</td>
<td>79 (44.1)</td>
</tr>
<tr>
<td>5-9</td>
<td>43 (24.0)</td>
</tr>
<tr>
<td>≥10</td>
<td>27 (15.1)</td>
</tr>
<tr>
<td>Type of occupation</td>
<td></td>
</tr>
<tr>
<td>Occupational and environmental medicine</td>
<td>11 (6.2)</td>
</tr>
<tr>
<td>Nursing</td>
<td>75 (41.9)</td>
</tr>
<tr>
<td>Physical therapy</td>
<td>49 (27.4)</td>
</tr>
<tr>
<td>Counseling psychologist</td>
<td>16 (8.9)</td>
</tr>
<tr>
<td>Industrial hygiene safety engineering</td>
<td>23 (12.8)</td>
</tr>
<tr>
<td>Other</td>
<td>5 (2.8)</td>
</tr>
</tbody>
</table>

Reliability and Validity Analysis

The results of the reliability and exploratory factor analyses are shown in Table 4. Cronbach alpha values greater than .60 and .90 are generally considered acceptable and highly reliable, respectively. Cronbach alpha values for all variables, excluding perceived risk (.69), were within the recommended range (>0.70), and thus the reliability of the main variables was considered acceptable. The analysis was performed without deleting items because none of the items impaired reliability.

The Kaiser Meyer Olkin statistic was 0.90 and the result of the Barlett test was Chi square\(_{174}=819.66\) (\(P<.001\)). Cronbach alpha values for all variables, excluding perceived risk (.69), were within the recommended range (>0.70), and thus the reliability of the main variables was considered acceptable. The analysis was performed without deleting items because none of the items impaired reliability.

The analysis was performed without additional adjustment.

The fit indices for the research model were as follows: Chi square\(_{174}=819.66\) (\(P<.001\)), goodness-of-fit index=0.91, root mean square residual=0.04, root mean square error of approximation=0.07, normed fit index=0.94, relative fit index=0.93, incremental fit index=0.96, comparative fit index=0.95, Tucker-Lewis index=0.95, and adjusted goodness-of-fit-index=0.88. The adjusted goodness-of-fit-index did not meet the criteria, but the overall model fit was satisfactory, and the other indices met the criteria. The results of confirmatory factor analysis (Table 5) showed that the paths to the observed variables were significant (\(P<.001\)) for all latent variables. The average variance extracted was >0.50 and the construct reliability was >0.70; therefore, convergent validity was demonstrated. In addition, the discriminant validity ensured
that the average variance extracted was higher than the square value of the correlation coefficient (Table 6).

### Table 4. Reliability and exploratory factor analysis.

<table>
<thead>
<tr>
<th>Variable (loading)</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
<th>Factor 6</th>
<th>Cronbach alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
</tr>
<tr>
<td>BI1</td>
<td>0.93</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.04</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>BI2</td>
<td>0.88</td>
<td>0.05</td>
<td>0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>BI3</td>
<td>0.80</td>
<td>0.01</td>
<td>0.06</td>
<td>0.05</td>
<td>0.06</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>EE(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.96</td>
</tr>
<tr>
<td>EE1</td>
<td>0.01</td>
<td>0.96</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.03</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>EE2</td>
<td>-0.004</td>
<td>0.94</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>EE3</td>
<td>-0.02</td>
<td>0.89</td>
<td>0.02</td>
<td>-0.02</td>
<td>0.09</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>EE4</td>
<td>0.04</td>
<td>0.83</td>
<td>0.01</td>
<td>0.06</td>
<td>-0.02</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>SI(^c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
</tr>
<tr>
<td>SI1</td>
<td>-0.03</td>
<td>0.04</td>
<td>0.92</td>
<td>-0.01</td>
<td>-0.02</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>SI2</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.89</td>
<td>-0.02</td>
<td>0.02</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>SI3</td>
<td>0.08</td>
<td>-0.04</td>
<td>0.81</td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>SI4</td>
<td>0.02</td>
<td>0.06</td>
<td>0.76</td>
<td>0.05</td>
<td>-0.03</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>PE(^d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.79</td>
</tr>
<tr>
<td>PE1</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.07</td>
<td>0.80</td>
<td>-0.01</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>PE2</td>
<td>-0.14</td>
<td>-0.02</td>
<td>0.05</td>
<td>0.74</td>
<td>0.05</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>PE3</td>
<td>0.16</td>
<td>0.05</td>
<td>0.02</td>
<td>0.67</td>
<td>-0.03</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>FC(^e)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.87</td>
</tr>
<tr>
<td>FC1</td>
<td>-0.07</td>
<td>0.07</td>
<td>-0.05</td>
<td>.001</td>
<td>0.88</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>FC2</td>
<td>0.08</td>
<td>-0.04</td>
<td>-0.07</td>
<td>0.01</td>
<td>0.82</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>FC3</td>
<td>0.08</td>
<td>0.05</td>
<td>0.07</td>
<td>0.01</td>
<td>0.68</td>
<td>-0.04</td>
<td></td>
</tr>
<tr>
<td>FC4</td>
<td>0.01</td>
<td>0.03</td>
<td>0.17</td>
<td>0.03</td>
<td>0.62</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>PR(^f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.69</td>
</tr>
<tr>
<td>PR1</td>
<td>0.09</td>
<td>-0.003</td>
<td>0.06</td>
<td>0.003</td>
<td>-0.09</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>PR2</td>
<td>-0.05</td>
<td>-0.06</td>
<td>-0.02</td>
<td>0.04</td>
<td>-0.03</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>PR3</td>
<td>-0.05</td>
<td>0.02</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.08</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>8.03</td>
<td>1.97</td>
<td>1.41</td>
<td>1.37</td>
<td>1.11</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Variance (%)</td>
<td>38.25</td>
<td>9.37</td>
<td>6.74</td>
<td>6.50</td>
<td>5.29</td>
<td>4.49</td>
<td></td>
</tr>
<tr>
<td>Cumulative variance (%)</td>
<td>38.25</td>
<td>47.61</td>
<td>54.35</td>
<td>60.85</td>
<td>66.15</td>
<td>70.63</td>
<td>.87</td>
</tr>
</tbody>
</table>

\(^a\)BI: behavioral intention.  
\(^b\)EE: effort expectancy.  
\(^c\)SI: social influence.  
\(^d\)PE: performance expectancy.  
\(^e\)FC: facilitating conditions.  
\(^f\)PR: perceived risk.
<table>
<thead>
<tr>
<th>Latent and observed variables</th>
<th>B</th>
<th>beta</th>
<th>SE</th>
<th>Critical ratio</th>
<th>P value</th>
<th>AVE&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CR&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BI</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.90</td>
<td>0.96</td>
</tr>
<tr>
<td>BI1</td>
<td>1.07</td>
<td>.96</td>
<td>0.02</td>
<td>47.39</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI2</td>
<td>1.06</td>
<td>.96</td>
<td>0.02</td>
<td>47.77</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI3</td>
<td>1.00</td>
<td>.89</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EE</strong>&lt;sup&gt;e&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
<td>0.96</td>
</tr>
<tr>
<td>EE1</td>
<td>0.90</td>
<td>.86</td>
<td>0.02</td>
<td>41.82</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE2</td>
<td>1.04</td>
<td>.92</td>
<td>0.02</td>
<td>54.17</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE3</td>
<td>1.00</td>
<td>.96</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE4</td>
<td>0.99</td>
<td>.93</td>
<td>0.02</td>
<td>56.61</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SI</strong>&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.75</td>
<td>0.92</td>
</tr>
<tr>
<td>SI1</td>
<td>1.01</td>
<td>.90</td>
<td>0.03</td>
<td>38.50</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI2</td>
<td>0.98</td>
<td>.88</td>
<td>0.03</td>
<td>36.89</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI3</td>
<td>1.00</td>
<td>.89</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI4</td>
<td>0.92</td>
<td>.78</td>
<td>0.03</td>
<td>29.17</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PE</strong>&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.72</td>
<td>0.88</td>
</tr>
<tr>
<td>PE1</td>
<td>0.89</td>
<td>.77</td>
<td>0.05</td>
<td>19.19</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2</td>
<td>0.87</td>
<td>.67</td>
<td>0.05</td>
<td>17.68</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE3</td>
<td>1.00</td>
<td>.82</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FC</strong>&lt;sup&gt;h&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.65</td>
<td>0.88</td>
</tr>
<tr>
<td>FC1</td>
<td>0.99</td>
<td>.77</td>
<td>0.04</td>
<td>24.14</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC2</td>
<td>1.12</td>
<td>.85</td>
<td>0.04</td>
<td>26.80</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC3</td>
<td>1.00</td>
<td>.80</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC4</td>
<td>0.90</td>
<td>.74</td>
<td>0.04</td>
<td>22.76</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PR</strong>&lt;sup&gt;i&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td>PR1</td>
<td>1.25</td>
<td>.76</td>
<td>0.10</td>
<td>12.30</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR2</td>
<td>1.27</td>
<td>.74</td>
<td>0.10</td>
<td>12.29</td>
<td>&lt;.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PR3</td>
<td>1.00</td>
<td>.50</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>AVE: average variance extracted.
<sup>b</sup>CR: composite reliability.
<sup>c</sup>BI: behavioral intention.
<sup>d</sup>N/A: not applicable.
<sup>e</sup>EE: effort expectancy.
<sup>f</sup>SI: social influence.
<sup>g</sup>PE: performance expectancy.
<sup>h</sup>FC: facilitating conditions.
<sup>i</sup>PR: perceived risk.
Hypothesis Testing

The results of Pearson correlation analysis showed that all dependent variables and the independent variable in the research model were significantly correlated (Table 7). In particular, perceived risk was negatively correlated with all other variables.

We performed multiple linear regression analysis to verify the effects of effort expectancy, social influence, performance expectancy, facilitating conditions, and perceived risk on behavioral intention. The factor analysis results showed that the regression model was statistically significant ($F=194.96$, $P<.001$), as shown in Table 8. The explanatory power of the regression model was 53.1% ($R^2=0.53$, $adjR^2=0.53$). Moreover, no issues were observed with respect to the independence of residuals (D-W=2.04) or multicollinearity (variance inflation factor<10).

The regression coefficients showed that effort expectancy, social influence, performance expectancy, and facilitating conditions exerted significant positive effects on behavioral intention, whereas perceived risk exerted a significant negative effect on behavioral intention (Table 8). Therefore, the results supported H1, H2, H3, H4, and H5.

We performed multigroup path analysis to verify the moderating effects of gender and age (Table 9). In the male group, effort expectancy, social influence, and performance expectancy exerted significant positive effects on behavioral intention. In the female group, effort expectancy, social influence, and performance expectancy exerted significant positive effects on behavioral intention. Only the performance expectancy (critical ratio=-3.38, $P<.001$) showed statistically significant differences in the path between males and females.

In the younger group, effort expectancy, social influence, and performance expectancy exerted significant positive effects on behavioral intention. In the older group, effort expectancy, social influence, and performance expectancy exerted significant positive effects on behavioral intention. Only the performance expectancy (critical ratio=1.97) showed statistically significant differences in the path between younger and older respondents. Therefore, the results supported H6, but H7 and H8 were rejected.

The path analysis of workers and health experts showed that effort expectancy, social influence, performance expectancy, and facilitating conditions exerted significant positive effects on behavioral intention, whereas perceived risk exerted a significant negative effect on behavioral intention. Only the perceived risk (critical ratio=-2.24) showed statistically significant differences in the path between workers and health experts.
Table 7. Pearson correlation coefficients among dependent variables.

<table>
<thead>
<tr>
<th></th>
<th>BI</th>
<th>EE</th>
<th>SI</th>
<th>PE</th>
<th>FC</th>
<th>PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td>0.51(^g)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td>0.61(^g)</td>
<td>0.47(^g)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>0.33(^g)</td>
<td>0.20(^g)</td>
<td>0.28(^g)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC</td>
<td>0.42(^g)</td>
<td>0.67(^g)</td>
<td>0.28(^g)</td>
<td>0.21(^g)</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>-0.53(^g)</td>
<td>-0.42(^g)</td>
<td>-0.32(^g)</td>
<td>-0.31(^g)</td>
<td>-0.34(^g)</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\(^a\)BI: behavioral intention.  
\(^b\)EE: effort expectancy.  
\(^c\)SI: social influence.  
\(^d\)PE: performance expectancy.  
\(^e\)FC: facilitating conditions.  
\(^f\)PR: perceived risk.  
\(^g\)\(P < .01\).  

Table 8. Multiple linear regression analysis with behavioral intention as the dependent variable.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>B</th>
<th>SE</th>
<th>beta</th>
<th>ts85</th>
<th>(P) value</th>
<th>VIF(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>-6.9E-17</td>
<td>0.02</td>
<td>0.00</td>
<td>.03</td>
<td>&gt;.99</td>
<td>2.24</td>
</tr>
<tr>
<td>EE(^b)</td>
<td>0.08</td>
<td>0.04</td>
<td>.08</td>
<td>2.17</td>
<td>&lt;.001</td>
<td>1.34</td>
</tr>
<tr>
<td>SI(^c)</td>
<td>0.43</td>
<td>0.03</td>
<td>.43</td>
<td>15.73</td>
<td>&lt;.001</td>
<td>1.17</td>
</tr>
<tr>
<td>PE(^d)</td>
<td>0.07</td>
<td>0.03</td>
<td>.07</td>
<td>2.67</td>
<td>.008</td>
<td>1.81</td>
</tr>
<tr>
<td>FC(^e)</td>
<td>0.14</td>
<td>0.03</td>
<td>.13</td>
<td>4.30</td>
<td>&lt;.001</td>
<td>1.81</td>
</tr>
<tr>
<td>PR(^f)</td>
<td>-0.32</td>
<td>0.03</td>
<td>-.29</td>
<td>-10.99</td>
<td>&lt;.001</td>
<td>1.31</td>
</tr>
</tbody>
</table>

\(^a\)VIF: variable inflation factor.  
\(^b\)EE: effort expectancy.  
\(^c\)SI: social influence.  
\(^d\)PE: performance expectancy.  
\(^e\)FC: facilitating conditions.  
\(^f\)PR: perceived risk.
### Table 9. Multigroup path analysis.

<table>
<thead>
<tr>
<th>Path Description</th>
<th>Male/ Younger (&lt;39 years)/Workers</th>
<th>Female/Older (≥40 years)/Health experts</th>
<th>Critical ratio for difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>beta</td>
<td>SE</td>
</tr>
<tr>
<td>Gender (Male or Female)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE\textsuperscript{a} to BI\textsuperscript{b}</td>
<td>0.10</td>
<td>.09</td>
<td>0.07</td>
</tr>
<tr>
<td>SI\textsuperscript{c} to BI</td>
<td>0.38</td>
<td>.40</td>
<td>0.05</td>
</tr>
<tr>
<td>PE\textsuperscript{d} to BI</td>
<td>0.04</td>
<td>.06</td>
<td>0.07</td>
</tr>
<tr>
<td>Age (Younger or Older)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE to BI</td>
<td>0.07</td>
<td>.06</td>
<td>0.07</td>
</tr>
<tr>
<td>SI to BI</td>
<td>0.39</td>
<td>.39</td>
<td>0.05</td>
</tr>
<tr>
<td>PE to BI</td>
<td>0.13</td>
<td>.12</td>
<td>0.08</td>
</tr>
<tr>
<td>Group (Workers or Health experts)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE to BI</td>
<td>0.08</td>
<td>.08</td>
<td>0.05</td>
</tr>
<tr>
<td>SI to BI</td>
<td>0.38</td>
<td>.40</td>
<td>0.03</td>
</tr>
<tr>
<td>PE to BI</td>
<td>0.11</td>
<td>.13</td>
<td>0.05</td>
</tr>
<tr>
<td>FC\textsuperscript{e} to BI</td>
<td>0.13</td>
<td>.12</td>
<td>0.05</td>
</tr>
<tr>
<td>PR\textsuperscript{f} to BI</td>
<td>-0.18</td>
<td>-.20</td>
<td>0.07</td>
</tr>
</tbody>
</table>

\textsuperscript{a}EE: effort expectancy.
\textsuperscript{b}BI: behavioral intention.
\textsuperscript{c}SI: social influence.
\textsuperscript{d}PE: performance expectancy.
\textsuperscript{e}FC: facilitating conditions.
\textsuperscript{f}PR: perceived risk.

### Differences in Perception between Workers and Health Experts

The results of the independent samples t tests to analyze differences in perception of the main variables between workers and health experts are summarized in Table 10. Workers’ mean scores for the main variables were higher relative to those of health experts for all variables except perceived risk. Moreover, behavioral intention, effort expectancy, social influence, performance expectancy, and perceived risk differed significantly between groups, whereas facilitating conditions did not.

### Table 10. Perception differences between groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Health experts, mean (SD)</th>
<th>Workers, mean (SD)</th>
<th>t\textsubscript{664}</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI\textsuperscript{a}</td>
<td>3.44 (0.97)</td>
<td>3.72 (0.82)</td>
<td>-3.58</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>EE\textsuperscript{b}</td>
<td>3.43 (0.79)</td>
<td>3.70 (0.83)</td>
<td>-3.90</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SI\textsuperscript{c}</td>
<td>2.92 (0.92)</td>
<td>3.34 (0.86)</td>
<td>-5.69</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>PE\textsuperscript{d}</td>
<td>3.90 (0.58)</td>
<td>4.01 (0.60)</td>
<td>-2.20</td>
<td>.03</td>
</tr>
<tr>
<td>FC\textsuperscript{e}</td>
<td>3.59 (0.73)</td>
<td>3.63 (0.84)</td>
<td>-0.69</td>
<td>.49</td>
</tr>
<tr>
<td>PR\textsuperscript{f}</td>
<td>2.84 (0.69)</td>
<td>2.56 (0.69)</td>
<td>4.85</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

\textsuperscript{a}BI: behavioral intention.
\textsuperscript{b}EE: effort expectancy.
\textsuperscript{c}SI: social influence.
\textsuperscript{d}PE: performance expectancy.
\textsuperscript{e}FC: facilitating conditions.
\textsuperscript{f}PR: perceived risk.
Discussion

The purpose of this study was to analyze the factors influencing users’ behavioral intention to use a PHR app and identify differences in perception of the main variables to inform the development of personalized health care services for workers. We developed a research model that added perceived risk to the UTAUT, which is a representative theory that explains information technology acceptance, and conducted an empirical study involving health experts and workers who visited 21 workers’ health centers. After receiving ethical approval from the Korea National Institute for Bioethics Policy, 1050 questionnaires were distributed over 7 weeks to 40 workers and 10 health experts from each workers’ health center; 900 completed questionnaires were collected. The number of respondents in the analysis was 866 (687 workers and 179 health experts).

Performance expectancy exerted significant positive effects on behavioral intention to use the PHR app. These results are consistent with those of previous studies [24,51,56-58,62,63]. Lee et al [56] demonstrated that performance expectancy was an important determinant of app use behavioral intention between college students and workers. Liu et al [57] showed that people outside the normal range of body mass index show the closest relation to performance expectancy with a fitness app’s behavioral intention. Mattila et al [18] showed that workers’ requirements for personal health technologies were simplicity, integration with everyday life, and clear feedback. To increase the behavioral intention of workers on the PHR app, it is necessary to both efficiently represent the causal relationship between physiological conditions associated with the collected data, and provide functions that can help workers’ health conditions through immediate feedback from health experts.

Effort expectancy exerted significant positive effects on behavioral intention to use the PHR app. These results are consistent those of previous studies [51,56,57,59,62,79]. Koivumäki et al [59] showed that effort expectancy was an important factor in consumers’ behavioral intention for electronic health services using personal data. Wang et al [79] showed that effort expectancy affects the intention to use a health care app, regardless of whether or not it is used. Ensuring the ease of use of the PHR app was a top priority for developers and planners because health care services are used at various ages and in various groups; therefore, the design and use of the method must be very intuitive. No matter how much the PHR app offers, if the interface is complex, users will stop using it.

Social influence exerted significant positive effects on behavioral intention to use the PHR app. These results are also consistent with previous studies [24,53,57,58,62,63,83,93]. Tan et al [93] showed that social influence significantly affected the intention to use personal digital assistance devices among medical professionals. Homburg et al [104] showed that bosses’ and colleagues’ opinions affect the intention of subordinates to adopt new technologies. Balk-Møller et al [105] showed that social function such as a peer challenge was used for a longer time to improve workplace health promotion than other app functions. Social influence demonstrates an important role between individuals and groups in the social ecology of workers, such as colleagues, employers, and health care professionals.

Employers can influence workers’ behavioral intention by providing an organizational culture that facilitates using the PHR app through health management policies that promote a healthy workplace environment. Workers exerted a synergistic effect on health promotion practices if employers and health managers played an active role in workplace health promotion programs [106]. Considering that social influence exerted the strongest effect on behavioral intention to use the PHR app, the app should include functions that enable interaction between colleagues or health experts.

In addition, workers’ behavioral intention increased when the app was linked to workplace health promotion activities or management direction of community institutions such as workers’ health centers. Workers’ health centers have conducted community institution activities such as occupational health care services for workers, on-site consultation services for the workplace, and establishing cooperative systems through networking among various institutions in the community [107]. Moreover, workers’ health centers provide personalized health care services to individual workers through an understanding of workers’ areas and the characteristics of industrial parks and working environments [108]. To continuously develop workers’ health centers, it is necessary to build practices that systematically collect information regarding workers’ medical checkups and harmful factors in the workplace and apply this information to follow-up management of the centers [107]. Under these circumstances, data collected via the PHR app can be used for personalized health care services and follow-up management of workers’ health centers.

Facilitating conditions exerted significant positive effects on behavioral intention to use the PHR app. These results are consistent with previous studies [53,58,63,83]. Stieglitz and Brochmann [109] proposed that facilitating conditions can be divided into material support (eg, incentives) and nonmaterial support (eg, training). It is difficult to retain workers who are not interested in continuous workplace health promotion participation. If the program emphasizes external motivations such as incentives rather than cycles through which personal motivation can be generated, the degree of participation will initially increase but cannot be sustained in the long run. It is important to configure an infrastructure that can support participants. Health managers in the workplace or clinicians in the clinic can run various health care programs using the PHR app and select individuals with risk factors to help them learn about healthy lifestyles. Employers can influence workers’ behavioral intention by establishing usage training, technical support teams, and organizational policies for the PHR app, along with wearable devices. Iron Mountain, a records and data management company, is running LiveWell [110], a worker health care program that uses wearable devices and apps. The program motivated workers to improve their health through policies that provide workers with various tasks and paid cash points are given to those who complete the tasks.

Perceived risk exerted a significant negative effect on behavioral intention to use the PHR app. The results showed that a greater
possibility of loss from using the PHR app was associated with lower behavioral intention. This reflected users’ concerns regarding the potential risks of using the PHR app, including potential personal information breaches. These results are consistent with previous studies [66,70,93], in which perceived risk affected the intention to use a particular technology. Choia et al [19] showed that potential risk factors such as workers’ personal health information and personal location collected from wearable devices affect workers’ intention to adopt the technology. Construction managers have also stressed the importance of addressing privacy issues before introducing wearable devices. Dawson et al [47] emphasized that workers are concerned about trust and confidentiality when accepting PHRs. Burkhard et al [48] noted that workers have concerns about completeness and accuracy as well as the privacy and security of PHRs. Health care technologies in the workplace enable data-based human resource management through health risk assessment for each worker. The employer may prevent disease by following a worker’s condition, but the worker may be concerned about the personal disadvantages of employers obtaining and tracking personal health information. Therefore, employers need to consider the workers’ perspective before introducing a PHR app in the workplace. It is important to gain the trust of workers and protect personal information through transparency of goals and procedures. Additionally, employers need to consider that (1) the purpose of the PHR app should be to facilitate a healthy culture within the workplace through workers’ health promotion and disease prevention, rather than to facilitate personnel management; (2) transparency of procedures, including the purpose and utilization, should be maintained and disclosed to PHR app users; and (3) the PHR app should protect privacy with user-centered design and operation.

To protect and maintain workers’ health, Korean workplaces should provide medical checkups for workers at institutions that have been designated by the Ministry of Employment and Labor or under the national health insurance law. The results of medical checkups are confidential and include not only personal information but also family history, lifestyle habits, previous history, and disease status. This information requires higher levels of protection than that required for general personal information. The results are sent directly to individual workers to strengthen personal information protection; however, this is inconvenient for health managers in the performance of their duties, because workers who lose or do not present their medical checkup results during consultations with health managers do not receive accurate consultations [111]. The PHR app can be used to collect and manage these results and share health information with specific medical personnel with consent. Moreover, it is necessary to develop security devices and specify security responsibilities for each step of medical data processing to increase the behavioral intention to use the PHR app.

The association between performance expectancy and behavioral intention of the PHR app was moderated by gender and age, where performance expectancy was higher in female and younger participants. These results are consistent with previous studies [62], Wang et al [112] showed that women have a higher preference for health care apps than men. Guo et al [102] argued that younger individuals showed a positive attitude toward the use of new technology, whereas older individuals were often slower to acquire technology. Adas [113] reported that men were more interested in technology than women, but Fitzgerald et al [114] argued that women are more interested in health status and are therefore more likely to seek medical advice and preventive care than men. The present study showed that gender and age did not moderate the effects of effort expectancy, social influence, or behavioral intention. These results are consistent with previous studies [62,99]. Gender and age are important factors in the health care environment, but there is no strong evidence identifying their specific roles [99]. Therefore, future research should focus on moderating variables such as gender, age, educational background, and app use experience.

In addition, behavioral intention, effort expectancy, social influence, performance expectancy, and perceived risk differed between workers and health experts; workers’ mean scores for the main variables were higher than those of health experts. Facilitating conditions scores did not differ between the two groups. Previous studies have identified differences in patients’ and providers’ perceptions, attitudes, and preferences regarding health care technologies [115], including PHRs [116]. The present study identified differences in perspective between consumers and providers regarding PHRs. PHRs are similar to EHRs in that they collect and manage individual health-related information in one place, but can be distinguished from EHRs in that individuals have ownership or control of the information. That is, providers can obtain information from PHRs only when authorized through access controls set by the consumer. According to related studies, patients with a strong need for clinical services who had chronic diseases with complications were highly likely to use PHRs [117-119], and most empirical studies have shown that patients were highly satisfied with their PHRs [119-121]. In contrast, some studies have shown that doctors were less likely to expect benefits from their patients’ use of PHRs and were concerned about the impact of PHR use on their workloads [122,123]. However, the workload burden resulting from PHR use was found to be lower than expected, and medical personnel were generally satisfied with PHRs [120,121,124]. Since PHRs are currently in the introduction stage in Korea, most health care professionals, including health experts at workers’ health centers, have no practical PHR experience. As in previous studies, concerns about increased workload, record accuracy, and the negative impact on patients from information disclosure are judged to have affected health experts’ relatively low confidence in PHRs. PHR apps could help health care providers make decisions and provide information based on consumers’ health records; however, they should only be implemented after conducting sufficient research examining necessary information collection and functions to ensure a balance between providers’ and consumers’ needs.

This study is the first to examine the factors influencing behavioral intention to use a PHR app in the field of occupational safety and health in Korea. Most studies have focused on the intention of patients, the elderly, and providers to use EHRs, health care devices, and telemedicine services; however, few have analyzed the intention to use PHR apps for workplace health promotion. This study is meaningful in that
it reports on workers’ and health experts’ acceptance of interconnected PHR apps to improve workplace health promotion, but it also has some limitations. For example, the study included workers who visited workers’ health centers; therefore, data were collected mainly from workplaces with fewer than 50 employees. In addition, the health experts offering health care services at workers’ health centers are limited to workers in workplaces with fewer than 50 employees, and these small-scale employers are not obliged to appoint health managers. Consequently, the actual work performed by health managers may differ from that of workers’ health center health experts. To derive more generalizable research results, future research should include workers and health managers from different sized workplaces. In addition, the study included only basic participant characteristics such as gender and age. Future studies should examine behavioral intention to use PHR apps according to users’ health status, disease, experience, and working environments. Moreover, PHRs are currently in the introduction stage in Korea, and there has been minimal scholarly debate regarding the use of workers’ PHRs. The current results could change according to the purpose and function of PHRs. Therefore, future research should examine the functions and application range of PHR apps for workers and health managers. Further, this study focused on acceptance of the PHR app for workplace health promotion through a research model that added perceived risk to the UTAUT, but it is also necessary to analyze behavioral changes in health promotion facilitated by PHR app use. Therefore, future research will analyze changes in workers’ health promotion behavior associated with workplace PHR app services by applying the health belief model [125], a representative theory that explains changes in health behavior. Future study will also analyze workers’ usage logs collected through the service operation as well as participants’ lifestyle changes and risk factor changes associated with metabolic syndrome and service satisfaction.

Acknowledgments
This work was supported by the Creative Industrial Technology Development Program (20002708, development and commercialization of the personalized healthcare service for employees based on the PHR platform), which is funded by the Ministry of Trade, Industry and Energy in Korea. We would like to thank the Korea Occupational Safety and Health Agency and the 21 workers’ health centers nationwide for their help in data collection.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Overview of the instrument.
[DOCX File, 29 KB - mhealth_v8f6e16723_app1.docx ]

References


45. Kaebler D, Pan EC. The value of personal health record (PHR) systems. AMIA Annu Symp Proc 2008 Nov 06:343-347 [FREE Full text] [Medline: 18999276]


110. IRON MOUNTAIN: Live Well. URL: https://ironmountain.mywellmetrics.com/ [accessed 2020-02-25]


Abbreviations

- EHR: electronic health record
- FHIR: fast healthcare interoperability resource
- HL7: Health Level Seven
- mHealth: mobile health
- PHR: personal health record
- TAM: technology acceptance model
- UTAUT: unified theory of acceptance and use of technology
Experiences of African American Breast Cancer Survivors Using Digital Scales and Activity Trackers in a Weight Gain Prevention Intervention: Qualitative Study

Julianne M Power1, MS; Deborah F Tate1,2,3, PhD; Carmina G Valle2,3, PhD

1Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
2Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
3Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Julianne M Power, MS
Department of Health Behavior
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
302 Rosenau Hall, CB #7440
135 Dauer Drive
Chapel Hill, NC, 27514
United States
Phone: 1 9199665852 ext 2
Email: julma@live.unc.edu

Abstract

Background: The use of digital tools to promote daily self-weighing and daily activity tracking may be a promising strategy for weight control among African American breast cancer survivors (AABCS). There have been no studies exploring the acceptability and feasibility of using digital tools for weight control or qualitative studies characterizing perceptions of daily self-weighing and daily activity tracking among AABCS.

Objective: This study aimed to explore the subjective experiences of daily self-weighing and daily activity tracking using digital tools, including wireless scales and activity trackers, in a sample of AABCS participating in two technology-based weight gain prevention interventions over 6 months.

Methods: Semistructured interviews (N=21) were conducted in person or over the phone, were audio recorded, and then transcribed verbatim. Each transcript was read to identify key themes and develop a codebook. Each transcript was coded using Atlas.ti software, and code outputs were used to identify overarching themes and patterns in the data.

Results: On average, participants were 52.6 (SD 8.3) years of age, with obesity at baseline (BMI 33.1 kg/m², SD 5.9), and weighed on 123.4 (SD 48.0) days out of the 168 days (73.5%) in the study period. Women tended to attribute their weight gain to cancer treatment and framed program benefits in terms of improved quality of life and perceptions of prolonging their survival following treatment. Using the smart scale for daily self-weighing was viewed as the tool by which participants could control their weight and improve their health and well-being posttreatment. The activity tracker increased awareness of physical activity and motivated participants to be more active.

Conclusions: Participants reported positive experiences and benefits from daily self-weighing and daily activity tracking. Findings suggest that daily self-weighing and daily activity tracking using digital tools are well-received, acceptable, and feasible intervention strategies for AABCS in the context of posttreatment weight management.

(JMIR Mhealth Uhealth 2020;8(6):e16059) doi:10.2196/16059

KEYWORDS
African American; cancer survivors; digital tools; weight gain prevention; qualitative

https://mhealth.jmir.org/2020/6/e16059
Introduction

Background

Weight gain after breast cancer diagnosis has been associated with disease recurrence [1], increased risk of mortality [2], and other adverse and long-term effects [3]. African American women experience greater weight gain following breast cancer diagnosis [4] and worse survival outcomes compared with other breast cancer survivors [5,6]. A qualitative study found that African American breast cancer survivors (AABCS) viewed weight gain following cancer diagnosis as a significant stressor [7]. Postdiagnosis weight gain may be an important characteristic of breast cancer survivorship among African American women [7]; hence, weight gain prevention efforts should be prioritized for this high-risk population of survivors. Although the feasibility and safety of weight control interventions have been demonstrated for breast cancer survivors [8], African American women are underrepresented in these studies [9,10]. In a recent review, only 8 weight control studies were identified specifically for AABCS [11]. Few studies have incorporated web-based or mobile technology, despite demonstrating efficacy for weight management and the potential to improve reach among racial and ethnic minority populations [12]. Only one intervention encouraged self-monitoring of weight and physical activity using digital tools [13].

Self-monitoring is a key component of behavioral weight control programs [14], and more frequent self-monitoring of body weight and physical activity has been associated with weight loss [14,15]. One possible explanation comes from social cognitive theory (SCT), which posits that individual, social, and environmental factors interact in a reciprocal manner to explain behavior [16]. According to SCT, individuals cannot self-regulate their behavior if they do not understand the conditions under which the behavior occurs or the consequences of the behavior. Therefore, successful self-regulation depends upon consistent self-monitoring done in temporal proximity to the target behavior [16]. Digital tools, such as wireless scales and activity trackers, promote adherence to self-monitoring, which has traditionally been burdensome and diminishes over time [17]. These tools allow for daily self-monitoring of weight and physical activity in the context of weight control interventions, and enable real-time data collection and delivery of tailored feedback based on objectively monitored data.

Daily self-weighing is an effective strategy for weight control that has been associated with adoption of weight control behaviors, such as reduced calorie intake and increased caloric expenditure [18]. In the only study of weight gain prevention that promoted daily self-weighing among AABCS to date, the number of self-weighing days was significantly associated with weight change [13,19], and nonadherence to daily self-weighing and daily activity tracking was associated with weight gain [19]. Concerns remain about whether frequent self-weighing might produce negative psychological consequences [20]. A recent study found no negative psychological effects of daily self-weighing in the context of a weight gain prevention intervention [21]. However, participants were young adults aged between 18 and 35 years and were predominantly white; thus, information from more diverse populations is needed [21]. The use of digital tools to promote daily self-weighing and daily activity tracking may be a promising strategy for weight control among AABCS. However, there have been no studies exploring the acceptability and feasibility of using digital tools for weight control or qualitative studies characterizing perceptions of daily self-weighing and daily activity tracking among AABCS.

Objectives

Thus, this study aimed to explore the subjective experiences of daily self-weighing and daily activity tracking using digital tools, including wireless scales and activity trackers, in a sample of AABCS participating in the Weighing Every day for Love of Life and Body (WELL Body) program, a pilot randomized controlled trial that examined two technology-based weight gain prevention interventions for AABCS over 6 months [13]. To our knowledge, this is the first qualitative study of daily self-weighing.

Methods

Participants and Parent Study Design

To be eligible, participants had to be (1) female, (2) aged ≥18 years, (3) African American or black, (4) diagnosed with breast cancer in the last 10 years, (5) completed cancer treatment, and (6) showed no signs of progressive disease or a second primary cancer. Participants (N=35) were randomized to 1 of 3 groups: (1) daily self-weighing intervention (n=13); (2) daily self-weighing+ intervention (included daily activity tracking; n=11); or (3) delayed intervention control (n=11). The 6-month intervention focused on preventing weight gain via self-regulation of diet and physical activity, and daily self-weighing was promoted as the primary self-monitoring strategy. All participants were given a Bluetooth and Wi-Fi–enabled wireless scale (Withings Wireless Scale-30) [22] for use during the intervention; participants in the daily self-weighing+ intervention were also given an activity tracker (Withings Pulse) [23]. Participants had access to a website/mobile app with weight graphs and received weekly lessons and tailored feedback emails based on weight with or without activity data, depending on the intervention group. The interventions encouraged participants to view their weight information as a daily tool and indicator of progress with their diet and physical activity behaviors. The intervention took place between January 2014 and June 2015 in Chapel Hill, North Carolina. The details of the results are described elsewhere [13]. The intervention showed potential for preventing weight gain in AABCS. Median weight loss after 6 months was 0.2% in the daily self-weighing intervention and 0.9% in the daily self-weighing+ intervention versus 0.2% gain in the delayed intervention control [13]. This study was approved by the University of North Carolina (UNC) Lineberger Comprehensive Cancer Center protocol review committee and the institutional review board (IRB) of the UNC at Chapel Hill (IRB study number 13-2898). All participants provided informed consent.

Qualitative Study Design and Data Collection

For the current qualitative study, all women randomized to the daily self-weighing or daily self-weighing+ intervention groups
(n=24) were invited to participate in qualitative interviews postintervention. The qualitative subsample consisted of 21 of 24 participants who completed an interview within a month after the final 6-month assessment. Women were interviewed by 3 female PhD students and were compensated US $30. Semistructured interviews included questions focused on experiences using the digital tools provided for the intervention (Table 1). Interviews were conducted in person or by speaker phone in a private room and were audio recorded using a digital recorder. Audio recordings were transcribed verbatim and reviewed for accuracy. Demographic information was collected via baseline web-based questionnaires. Objective height and weight data were collected in person at a baseline clinic assessment.

**Table 1.** Semistructured interview questions.

<table>
<thead>
<tr>
<th>Number</th>
<th>Questions(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How would you describe your experience using the different technology tools provided by WELL Body(^b)?</td>
</tr>
<tr>
<td>1a</td>
<td>What barriers/problems, if any, did you have using the smart scale?</td>
</tr>
<tr>
<td>1b</td>
<td>What, if anything, would have helped you to successfully use the smart scale?</td>
</tr>
<tr>
<td>1c</td>
<td>What barriers/problems, if any, did you have using the app on your phone?</td>
</tr>
<tr>
<td>1d</td>
<td>What, if anything, would have helped you to successfully use the app?</td>
</tr>
<tr>
<td>1e</td>
<td>What barriers/problems, if any, did you have using the website?</td>
</tr>
<tr>
<td>1f</td>
<td>What, if anything, would have helped you to successfully use the website?</td>
</tr>
<tr>
<td>1g</td>
<td>What barriers/problems, if any, did you have using the activity tracker?</td>
</tr>
<tr>
<td>1h</td>
<td>What, if anything, would have helped you to successfully use the activity tracker?</td>
</tr>
<tr>
<td>2</td>
<td>Before WELL Body, what other health technology tools did you use for healthy eating, weight, or physical activity?</td>
</tr>
<tr>
<td>3</td>
<td>How would you describe your commitment to using the technology tools provided by WELL Body?</td>
</tr>
<tr>
<td>3a</td>
<td>Is there other information about the technology tools that you would have liked to help you in terms of how to use them to improve your diet or physical activity?</td>
</tr>
<tr>
<td>4</td>
<td>Think back to the beginning of the program, how confident were you that you could successfully do the WELL Body program? On a scale of 0-10 (0=not at all confident and 10=very confident).</td>
</tr>
<tr>
<td>4a</td>
<td>Were there parts of the program you were confident about that you used more than others? Describe.</td>
</tr>
<tr>
<td>5</td>
<td>A key recommendation of the WELL Body program is to weigh yourself every day, and each week, your homework and feedback suggested you continue to weigh daily and use your tracker. How well do you feel you accomplished this goal?</td>
</tr>
<tr>
<td>5a</td>
<td>How did you feel about weighing yourself every day?</td>
</tr>
<tr>
<td>5b</td>
<td>How successful were you in using the tracker every day?</td>
</tr>
<tr>
<td>6</td>
<td>What activities did you do as part of the WELL Body program?</td>
</tr>
<tr>
<td>6a</td>
<td>What activities/components worked best or were most helpful? How so? What ones didn’t work out well? Why?</td>
</tr>
<tr>
<td>6b</td>
<td>What activities/components were least helpful? How so?</td>
</tr>
<tr>
<td>6c</td>
<td>How did you feel about the weekly feedback? What, if anything, would have improved this portion of the program?</td>
</tr>
<tr>
<td>7</td>
<td>What plans, if any, do you have for continuing to use the smart scale provided by the WELL Body program?</td>
</tr>
<tr>
<td>8</td>
<td>What benefits do you think that you, or others (eg, family), gained from your participation in WELL Body?</td>
</tr>
<tr>
<td>8a</td>
<td>From using the smart scale?</td>
</tr>
<tr>
<td>8b</td>
<td>From using the activity tracker?</td>
</tr>
<tr>
<td>9</td>
<td>How could WELL Body be re-designed to make it easier to understand/use/maintain?</td>
</tr>
<tr>
<td>10</td>
<td>What would encourage you and other African American breast cancer survivors to participate in a study like this one again?</td>
</tr>
<tr>
<td>11</td>
<td>What advice would you give to other African American breast cancer survivors considering using a smart scale?</td>
</tr>
<tr>
<td>12</td>
<td>Is there anything you would like to add about your experiences?</td>
</tr>
</tbody>
</table>

\(^a\)Italicized text represents main questions whereas nonitalicized text represents subquestions.

\(^b\)WELL Body: Weighing Every day for Love of Life and Body.

**Qualitative Data Analysis**

The primary author (JMP) performed qualitative data analysis, first reading each transcript to identify key themes and develop a codebook with deductive codes derived from the interview guide [24]. *Cancer experience* was included as an inductive code because over half of participants talked about their experience with cancer. Each transcript was coded using Atlas.ti.
software [25] to explore which intervention components participants talked about most frequently. For this study, the following codes were analyzed: daily weighing, smart scale, activity tracker, and cancer experiences. Code outputs were used to further explore the ideas contained within each code through descriptive analytic summaries. The analytic summaries consisted of a description of each participant’s comments related to each code. By examining the analytic summaries together, overarching themes were identified. Matrices were prepared to visualize the data and analyze patterns across participants [26]. Memo writing facilitated interpretation of the data throughout the analysis process [27]. Results from the semistructured interviews are described using illustrative quotes to highlight key findings, and participants are referred to using pseudonyms.

**Results**

**Overview**

On average, participants were 52.6 (SD 8.3) years of age, obese at baseline (BMI 33.1, SD 5.9), weighed on 123.4 (SD 48.0) days, and tracked physical activity on 130.3 (SD 52.4) days out of the 168 days (73.5% and 77.6%, respectively) in the study period (Table 2). There were no differences between participants who completed the interviews (n=21) and those who did not (n=3). Women tended to attribute their weight gain to cancer treatment and framed the benefits of WELL Body in terms of improved quality of life (QoL) and perceptions of prolonging their survival following cancer treatment. Several barriers and facilitators were identified to using the smart scale and activity tracker. Using the smart scale for daily self-weighing was viewed as the tool by which participants could control their weight and improve their health and well-being posttreatment. The activity tracker increased awareness of physical activity and motivated participants to be more active.

**Table 2.** Characteristics of intervention participants in WELL body (N=24).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Qualitative subsample (n=21)</th>
<th>Nonqualitative subsample (n=3)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>52.62 (8.27)</td>
<td>51.00 (9.17)</td>
<td>.66</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), mean (SD)</td>
<td>33.12 (5.92)</td>
<td>34.51 (4.49)</td>
<td>.97</td>
</tr>
<tr>
<td>Years since diagnosis, mean (SD)</td>
<td>3.38 (2.46)</td>
<td>1.33 (1.16)</td>
<td>.12</td>
</tr>
<tr>
<td>&lt;High school education level, n (%)</td>
<td>2 (9.5)</td>
<td>0 (0)</td>
<td>.07</td>
</tr>
<tr>
<td>Annual income &lt;US $60,000, n (%)</td>
<td>8 (47.1)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2 (66.7)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Married, n (%)</td>
<td>11 (52.4)</td>
<td>1 (33.3)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Premenopausal, n (%)</td>
<td>4 (19.0)</td>
<td>1 (33.3)</td>
<td>.52</td>
</tr>
<tr>
<td>Days weighed over 6 months, mean (SD)</td>
<td>123.38 (47.97)</td>
<td>89.00 (8.54)</td>
<td>.09</td>
</tr>
<tr>
<td>Days physical activity tracked over 6 months, mean (SD)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>135.30 (52.36)</td>
<td>80.00</td>
<td>.34</td>
</tr>
<tr>
<td>Percent weight change at 6 months, mean (SD)</td>
<td>-2.65 (5.15)</td>
<td>0.89 (0.77)</td>
<td>.11</td>
</tr>
</tbody>
</table>

<sup>a</sup>On the basis of Kruskal-Wallis tests for continuous variables and Fisher exact tests for categorical variables.

<sup>b</sup>n=17 (4 cases missing).

<sup>c</sup>Daily self-weighing+intervention group only; n=10 (qualitative subsample) and n=1 (nonqualitative subsample).

**Weight Gain Experiences During and After Cancer**

One-third of participants (n=7) talked about how treatment affected their bodies, resulting in weight gain. Janet, age 57 (1 year since diagnosis), portrayed her weight gain as a process from which she felt somewhat removed. She only became aware of her weight gain after the demands of treatment were over:

> For me it was only chemo, well not only chemo, but chemo and radiation and I was telling her I was sharing with friends you know I really did not notice the gain until after everything was over.

Participants described negative physical and psychological consequences of treatment, such as Carol, age 54 (2 years since diagnosis). Physical effects included limited mobility and fatigue. Psychological effects included negative body image and low self-esteem. These physical and psychological effects made it difficult for Carol to lose weight, which left her searching for a way to address her weight gain:

> You hit such a low after diagnosis and treatment, and sometimes you’re not exactly healthy with the way your body looks and feels after that. You’re looking for any opportunity to, umm, make yourself feel pretty, look better, or feel better about yourself, and the weight is a big thing. So I gained a lot of weight through treatment and after, and it’s harder to get it off, so, ’cause you can’t move like you were before and you don’t have the stamina.

Participants associated experiences of weight gain during and after cancer with a lack of control as well as negative physical and psychological consequences. Women were motivated to address weight gain resulting from treatment.
Benefits of the Weighing Every Day for Love of Life and Body Program

Most women (n=13) contextualized their experience in WELL Body within their broader experience of cancer. A total of 5 participants framed the main program benefits in terms of enhanced recovery from treatment. Lisa, age 36 (8 years since diagnosis), viewed WELL Body as a way to deal with the weight gain and body image issues that stemmed from treatment by addressing lifestyle factors that impact overall health, improving her QoL:

Um, being a breast cancer survivor, well for me, I don’t know why I got it. Um, and I guess the best thing to do is try and live a life where you feel good, and in order to feel good, you have to be as healthy as you can... And in order to be as healthy as you can, you have to eat better, you have to exercise, so just having that sense of feeling better, and knowing that there’s a program to help you, um, feel good, and be healthy.

These ideas were reflected in the comments of other participants. WELL Body allowed Elizabeth, age 45 (5 years since diagnosis), to exert control over her weight, which she believed would decrease the likelihood of cancer recurrence and increase her chances of survival:

I am very determined to survive for as long as I can...so I want to do anything that’s going to potentially help me to do that... But I do know there are a lot of African American breast cancer survivors—whether they’re overweight or whether they’re at their ideal weight, we all struggle with the “Why?” sometimes and I think we can target a certain portion of our lives that might have an impact on whether we get it or not.

Women viewed WELL Body as a way to address weight gain resulting from treatment and control lifestyle factors that contribute to cancer risk, such as diet and physical activity. WELL Body provided an opportunity for participants to transition from treatment to survivorship by addressing long-term side effects of treatment, including weight gain.

Barriers and Facilitators to Daily Self-Weighing

Daily self-weighing via the smart scale is an integral component of the WELL Body program. Participants described several barriers to daily self-weighing as well as corresponding facilitators, or factors that helped them overcome these barriers.

Initiation and Creation of a Routine

Initiating daily self-weighing was a barrier for many participants because it was viewed as an extra chore. However, creating a daily weighing routine reduced the perceived burden of daily self-weighing. Among the one-third of participants who talked about the importance of routine (n=7), the most common routine described was weighing first thing in the morning before showering. Patricia, age 48 (10 years since diagnosis), who weighed 5 days/week on average, described:

So I didn’t like the fact of starting weighing myself because it had to be a routine with me, because I had to put it into my daily morning, and you know some mornings I would forget, so, umm, as far as the weighing every day in the beginning, was a hassle but then once I put it into my routine it was good.

Weight Fluctuations and Increased Awareness

Another barrier to daily self-weighing was weight fluctuations. Over a third of participants (n=9) reported feeling discouraged by weight fluctuations. Helen, age 54 (1 year since diagnosis), who weighed 5 days/week on average, was very conscious of her weight and described feeling nervous about daily self-weighing when she expected her weight to be up:

Sometimes I dreaded getting on the scale, it’s like, “omigod, I’m gonna be up today.” It was a little emotional at times in just dealing with the “what is it gonna be today because I ate so and so yesterday,” to be honest with you, it was quite stressful sometimes.

Despite feeling discouraged at times, Helen reflected that by continuing to weigh daily, she became more comfortable seeing her weight fluctuate and associated these fluctuations with her diet and physical activity behaviors:

I think overall it was a positive experience because it really, especially the scale, gave me the insight into my daily routine and I am more conscious about what I eat, how much I eat, but I can say I’m truly more conscious and this study has helped me to do that... I have a better understanding of how my behavior impacted my eating habits and my activity habits, and how that really impact my daily weight.

Over half of the participants (n=14) talked about how daily self-weighing made them more aware of their diet and physical activity behaviors, which is consistent with the intervention messages participants received. Mary, age 65 (4 years since diagnosis), who weighed 6 days/week on average, described how her negative emotions toward the scale were neutralized once she realized that weight fluctuations provided her with useful information about her behaviors:

I don’t get uptight when I get on the scale anymore. That’s made a big difference and always it encourages me to weigh by just looking at it as information and just the way I look at my blood sugar... if my blood sugar goes up I don’t stop taking my blood sugar medicine, I evaluate what’s going on and I let the doctor know. So the same thing with the scale, I’m using it as a piece of health information and not, this study has really helped me not to negatively equate the scale with the negative connotation of myself.

Some participants even looked forward to stepping on the scale, such as Elizabeth, age 45 (5 years since diagnosis), who weighed 4 days/week on average. Daily self-weighing gave Elizabeth an awareness that she could control her weight. After her experience with cancer, which felt so out of her control, this realization was empowering:

I think sometimes we operate sort of in a black hole, it’s better not to know, and this forced me to know on a daily basis and so I got to the point where I really...
enjoyed that. I wanted to know where I was and how I could adjust those numbers and realizing that I’m completely in control for the most part, of how my weight and my BMI fluctuate. So, I think that’s what it was rooted in, just having control over a certain part of my life when most of your life often seems like you have no control over it...

Nonadopters

Although most participants (n=18) ultimately adopted daily self-weighing and weighed ≥4 days/week on average, 3 participants expressed a preference not to engage in daily self-weighing and weighed less often. Donna, age 54 (4 years since diagnosis), weighed 2 days/week on average, and described daily self-weighing as discouraging because it aroused almost a decade’s worth of frustration resulting from “out of control” feelings toward her weight. Donna’s perspective provides insight into how emotionally charged daily self-weighing can be, which is a significant barrier to adoption:

Sometimes when I’m doing exactly what I’m supposed to be doing it’s like the scale is fluctuating instead of it going [down] it’s like I’m gaining and so that frustration and just used to being at a certain weight... I feel like I’m out of control.

Daily self-weighing via the smart scale was viewed by most participants as a tool to control their weight and achieve a healthier lifestyle. However, it is important to consider how a person’s weight history might influence their desires and preferences for adopting daily self-weighing.

Barriers and Facilitators to Daily Activity Tracking

Participants in the daily self-weighing+ intervention (n=10) group who self-monitored physical activity generally had positive feedback about their experiences using the activity tracker. Barriers included forgetting or losing the device (n=6), having problems syncing the device (n=6), and wanting more instruction on how to use the device (n=3). Linda, age 50 (2 years since diagnosis), tracked physical activity 7 days/week on average, and expressed concern about not knowing whether her device was functioning properly during the intervention:

There were times when I wasn’t sure it was working and I didn’t really understand why...sometimes I think you lose something when you have to recharge it because you have to reboot it and all this stuff. I think there were some days that it didn’t even calculate all the steps that I had walked... I think that needs a little more instruction to come with that.

In total, 2 participants talked about how the tracker worked well in conjunction with the scale to provide a comprehensive picture of their weight loss progress. Facilitators included feeling motivated by the device to reach a daily activity goal (n=7). Helen, age 54 (1 year since diagnosis), tracked her physical activity 5 days/week on average, and described how the activity tracker kept her accountable by triggering her sense of responsibility to get as much activity as possible on the days she wore the device:

I don’t know if you’ve seen the commercials where the person’s conscience is talking to him; it served as that, a little buddy on my shoulder, keeping me in line and keeping me on track of what I’m doing, and kinda forcing me to think about how to maintain and even to improve...

Sandra, age 65 (2 years since diagnosis), tracked her activity 6 days/week on average and reflected that knowing she was accumulating steps over the course of the day alleviated her worry about exercising in general and may have increased the perceived accessibility of a physically active lifestyle:

This is when I realized that your movement should be throughout the day, not just one sporadic situation where you are really working out but then the rest of the day you are just sitting around... I think it’s more moving than... worrying about exercising, but getting up and moving.

Discussion

Principal Findings

The purpose of this study was to explore the subjective experiences of daily self-weighing and daily activity tracking using digital tools in a sample of AABCS participating in two technology-based weight gain prevention interventions. Women viewed daily self-weighing as a way to control their weight and improve their health and well-being posttreatment. Weight fluctuations were identified as a significant barrier to daily self-weighing; however, for most, continued daily self-weighing normalized these fluctuations, allowing participants to view weight fluctuations as objective health information. The activity tracker increased awareness of physical activity and motivated participants to be more active. These findings suggest that using digital tools to self-monitor weight and physical activity is an acceptable self-regulation strategy for weight management among AABCS. Daily self-monitoring and access to real-time data may have given women a greater sense of agency and motivated them to engage in weighing and exercise behaviors.

Experiences of weight gain during and after cancer were associated with perceived lack of control. A cross-sectional study of breast cancer survivors 5 years postdiagnosis and healthy women (N=328) showed that breast cancer survivors generally perceive the world as less controllable and more random compared with healthy women; however, survivors perceive the same control over their own daily lives as healthy women [28]. This may explain why women in this study viewed their weight gain during and after cancer as a process that was largely out of their control, but were motivated to take charge of their weight by participating in WELL Body. Women viewed addressing their weight gain as a way to transition to survivorship after the illness experience. Weight gain may be viewed as a persistent symptom of cancer and its treatment [7], and continuing symptoms of cancer have been associated with psychological distress among survivors [29]. Findings from this study indicate that posttreatment weight management efforts could enhance survivorship among AABCS by increasing perceptions of personal control and reducing psychological distress related to their cancer.
One of the main benefits of the program was that participants perceived that changing their behaviors and weight would prolong their survival, which may be related to the management of uncertainty about cancer recurrence. In a study examining beliefs about breast cancer recurrence risk reduction among AABCS, more than half of the 191 respondents believed that being overweight was associated with breast cancer recurrence, and one-third agreed that losing weight could prevent recurrence [30]. A qualitative study found that breast cancer survivors were motivated to join a lifestyle intervention by fear of cancer recurrence and a desire to know they had done everything to prevent recurrence [31]. Qualitative results from a lifestyle intervention indicate that breast cancer survivors gained a sense of control over their bodies and cancer recurrence by participating [32]. In this study, frequent weight and activity information may have given participants an increased sense of control over their bodies and facilitated a perceived reduction in risk of cancer recurrence, leading to perceptions of prolonging their posttreatment survival.

In total, 3 participants expressed a preference not to engage in daily self-weighing, suggesting that daily self-weighing can be emotionally charged depending on a participant’s weight history. However, other studies have found no adverse psychological effects associated with daily self-weighing [21,33]. Gorin et al [21] found that self-weighing frequency was not associated with depressive symptoms or binge eating, but was associated with improvements in health-related QoL. Another study by Wing et al [33] found that increases in frequency of self-weighing were associated with decreases in depressive symptoms and decreases in the probability of reporting binge eating episodes per month. These studies indicate that daily self-weighing is an important aspect of weight control and may mitigate depressive symptoms and disordered eating. However, more research is needed to determine whether and to what extent daily self-weighing has any adverse psychological effects among AABCS and the larger population of breast cancer survivors as a whole.

Daily activity tracking appeared to promote awareness of current activity levels and increase motivation to be more physically active among AABCS. Another qualitative study found that daily activity tracking increased motivation to be physically active and promoted awareness around sedentary time in patients with breast cancer [34]. Taken together, these findings indicate that daily activity tracking, in conjunction with goals and advice tailored for women with breast cancer, may be an effective way to promote a more active lifestyle throughout the cancer care continuum, which could improve QoL outcomes [3]. This study identified barriers to activity tracker use, including forgetting or losing the device, which is consistent with another qualitative study [34]. Providing more detailed instructions on using the device and ensuring the availability of technical assistance may help facilitate daily activity tracking. Capitalizing on digital tools that are already integrated into participants’ everyday lives may also address barriers to daily activity tracking. These findings support the feasibility and acceptability of using activity trackers in interventions for AABCS. Future interventions promoting daily activity tracking in women with breast cancer may benefit by including messaging to remind participants to wear their trackers every day.

Study Limitations
A key study limitation is that women voluntarily participated in the weight gain prevention interventions and may have been more motivated to address weight gain compared with other AABCS. Additionally, 3 participants did not complete qualitative interviews, so these findings may not be representative of the entire study population. Considering that 3 participants disliked and ultimately did not adopt daily self-weighing, future studies could identify different approaches to self-monitoring that are more acceptable for nonadopters. This study may have been strengthened by the inclusion of additional coders, which would have created opportunities to discuss coding disagreements and refine the coding system as well as incorporating multiple perspectives into the analytic process.

Conclusions
This study provides a qualitative perspective of the effects of daily self-weighing and daily activity tracking in AABCS, which has previously only been explored quantitatively. Participants expressed shared positive experiences related to daily self-weighing and daily activity tracking, such as improved QoL, perceptions of prolonging their survival following cancer treatment, and greater awareness surrounding physical activity. Common barriers included difficulty initiating new habits, whereas common facilitators included creating a routine around daily self-weighing and daily activity tracking. These findings and other quantitative data [13,19] suggest that daily self-weighing and daily activity tracking are acceptable and feasible for AABCS in the context of posttreatment weight management. Daily self-weighing in this study was accompanied by health education and tailored feedback to help women make sense of weight fluctuations by relating them back to diet and physical activity behaviors. Less is known about the impact of daily self-weighing among cancer survivors in the absence of a weight control program. Additionally, few studies have explored whether daily self-weighing and daily activity tracking are useful for African American women during breast cancer treatment. Future research might explore the use of these tools during treatment to further examine the impact of daily self-weighing and daily activity tracking among African American women and to determine if preventing weight gain during treatment or posttreatment is the more optimal time for interventions aiming to improve a sense of personal control and QoL.

Acknowledgments
The parent pilot randomized controlled trial (WELL Body) was funded by a Developmental Research Award from the University of North Carolina at Chapel Hill (UNC) Lineberger Comprehensive Cancer Center. The authors thank the UNC Health

https://mhealth.jmir.org/2020/6/e16059

Jmir MHealth UHealth 2020 | vol. 8 | iss. 6 | e16059 | p.112

(page number not for citation purposes)
Registry/Cancer Survivorship Cohort participants for their important contributions and the cancer survivors who participated in the study. We gratefully acknowledge the staff of the UNC Weight Research Program for their valuable support, including Melissa Cochran, Dr Melissa Crane, Dr Brooke T Nezami, Candice Alick, Loneke Blackman, and Rachel K Bordogna, who provided excellent research assistance.

Conflicts of Interest
DFT reports serving on the Scientific Advisory Board for WW, a company that delivers weight management interventions with digital technology.

References


Abbreviations

- **AABCS**: African American Breast Cancer Survivors
- **IRB**: institutional review board
- **QoL**: quality of life
- **SCT**: social cognitive theory
- **UNC**: University of North Carolina
- **WELL Body**: Weighing Every day for Love of Life and Body
Patients’ Measurement Priorities for Remote Measurement Technologies to Aid Chronic Health Conditions: Qualitative Analysis

Sara Simblett, PhD, DClinPsy; Faith Matcham, PhD; Hannah Curtis, BSc; Ben Greer, BSc; Ashley Polhemus, MSE; Jan Novák, MSc; Jose Ferrao, PhD; Peter Gamble, MBA; Matthew Hotopf, PhD, FRCPsych; Vaibhav Narayan, MBA, PhD; Til Wykes, PhD; Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) Consortium

Corresponding Author:
Sara Simblett, PhD, DClinPsy
Institute of Psychiatry, Psychology and Neuroscience
King’s College London
Henry Wellcome Building, Psychology Department, 2nd Floor
London
United Kingdom
Phone: 44 2078480762
Email: sara.simblett@kcl.ac.uk

Abstract

Background: Remote measurement technology (RMT), including the use of mobile phone apps and wearable devices, may provide the opportunity for real-world assessment and intervention that will streamline clinical input for years to come. In order to establish the benefits of this approach, we need to operationalize what is expected in terms of a successful measurement. We focused on three clinical long-term conditions where a novel case has been made for the benefits of RMT: major depressive disorder (MDD), multiple sclerosis (MS), and epilepsy.

Objective: The aim of this study was to conduct a consultation exercise on the clinical end point or outcome measurement priorities for RMT studies, drawing on the experiences of people with chronic health conditions.

Methods: A total of 24 participants (16/24 women, 67%), ranging from 28 to 65 years of age, with a diagnosis of one of three chronic health conditions—MDD, MS, or epilepsy—took part in six focus groups. A systematic thematic analysis was used to extract themes and subthemes of clinical end point or measurement priorities.

Results: The views of people with MDD, epilepsy, and MS differed. Each group highlighted unique measurements of importance, relevant to their specific needs. Although there was agreement that remote measurement could be useful for tracking symptoms of illness, some symptoms were specific to the individual groups. Measuring signs of wellness was discussed more by people with MDD than by people with MS and epilepsy. However, overlap did emerge when considering contextual factors, such as life events and availability of support (MDD and epilepsy) as well as ways of coping (epilepsy and MS).

Conclusions: This is a unique study that puts patients’ views at the forefront of the design of a clinical study employing novel digital resources. In all cases, measuring symptom severity is key; people want to know when their health is getting worse. Second, symptom severity needs to be placed into context. A holistic approach that, in some cases, considers signs of wellness as well as illness, should be the aim of studies employing RMT to understand the health of people with chronic conditions.

(JMIR Mhealth Uhealth 2020;8(6):e15086) doi:10.2196/15086

KEYWORDS qualitative analysis; patient involvement; remote measurement technology; mHealth
Introduction

It is estimated that by 2020, chronic health conditions will contribute to approximately 57% of the global burden of disease [1]. There is a need for innovative ways to support all these people in accessing clinical care and in managing their long-term conditions, in the context of limited resources. A case has been made for the use of mobile technology (eg, mobile phone apps and wearable technology) to provide real-world assessment and intervention that will both streamline clinical input and, where possible, promote independent self-management [2,3]. As an example, remote measurement technology (RMT) can gather data that may enable the early detection of worsening symptoms with the potential to offer rapid interventions. A recent systematic review identified an emergence of studies in this area [4]. In order to establish the benefits of using RMT, we need to clarify what is expected in terms of a successful outcome. The selection of outcomes measured is often determined by the interests of researchers, which may in part be driven by the availability of valid and reliable tools. More and more, people are advocating for the involvement of the people who receive or provide health services in translational research design [5]. The recent Academy of Medical Sciences’ report [6], Our data-driven future in healthcare: People and partnerships at the heart of health-related technologies, recommends that patients and the public should be active partners in agreeing on priorities for, and determining the acceptability of, data-driven technologies as part of an ongoing process.

The aim of this study was to conduct a consultation exercise on measurements of interest in RMT studies. We identified three clinical groups where there is a strong case for the potential benefits of RMT for measuring and managing recurrent and persistent chronic health conditions: major depressive disorder (MDD), a mood disorder characterized by a persistent feeling of sadness or a lack of interest in outside stimuli with a high risk of reoccurrence [7]; multiple sclerosis (MS), a relapsing or progressive demyelinating disease in which the insulating covers of nerve cells in the brain and spinal cord are damaged over time; and epilepsy, a long-term neurological condition causing frequent seizures (see Table 1). These clinical groups are varied and have different presenting characteristics, but they are all long-term, highly variable conditions that are costly to manage with largely unknown mechanisms precipitating relapse. Monitoring symptoms over time could assist with developing a better understanding of these relapse mechanisms and patterns of variability; this could potentially lead to the early identification of relapse or deterioration with the ability to intervene more quickly. Previous consultation exercises with people living with these three health conditions—MDD [8,9], MS [10], and epilepsy [11]—have identified overarching themes that are pertinent to the introduction of mobile technologies, including the importance of self-management, prevention or prediction of symptoms, and early intervention. None of these consultations so far have asked specifically about people’s views on what might be important to measure when implementing mobile technologies. This information is crucial for being able to design systems that engage users, under the assumption that measurement of meaningful information is necessary for sustained engagement [4]. The aim of this paper is to extend previous work and focus on the use of RMT.

Table 1. Case examples of the use of remote measurement technology.

<table>
<thead>
<tr>
<th>Health condition</th>
<th>Case example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major depressive disorder (MDD)</td>
<td>Symptom recall for people with MDD is frequently interrupted and biased by poor cognition and dysfunctional perceptions. Reliance on self-report measures alone leads to imprecise and inefficient estimations of effects in clinical trials. Mobile technology, including wearable sensors, may allow for more momentary and continuous assessment of factors associated with MDD (eg, reduced activity or change in speech patterns and other physiology). Signs of relapse may be able to be detected before a person is fully aware of their declining mood.</td>
</tr>
<tr>
<td>Multiple sclerosis (MS)</td>
<td>There is emerging evidence for the reliability and validity of mobility and gait assessment using wearable activity monitoring (ie, accelerometry) for modelling relapse in MS. Use of mobile sensors, combined with more frequent (eg, daily or weekly) self-reported outcomes to contextualize changes in activity, may provide early indicators of relapse that have not been detectable in the past.</td>
</tr>
<tr>
<td>Epilepsy</td>
<td>Routine electroencephalogram electrode technology for monitoring health state in epilepsy cannot be implemented for more than a few days at a time. There is scope to integrate mobile technology into clinical assessment that will allow collection of continuous data to track, and possibly predict, seizure occurrence as part of daily life. Other mobile sensors (eg, wearable heart rate and activity monitors) are being investigated as alternative, potentially less obtrusive, options.</td>
</tr>
</tbody>
</table>

Methods

Design

A qualitative approach using a thematic analysis was employed to elicit views on measurement priorities from service users. Themes and subthemes were identified following grounded-theory methods.

Context

Researcher Characteristics

Six focus groups were facilitated by two women—a clinical psychologist and a health psychologist—who were not involved in the participants’ clinical care.

Participant Characteristics

Participants were identified by convenience sampling and their eligibility to participate. Participants were included if they were...
over the age of 18 and had received a diagnosis of MS, epilepsy, or MDD (within the past 2 years for MDD). People with MS and epilepsy were recruited through third-sector organizations (ie, the MS Society and Epilepsy Action) and local clinics; people with MDD were recruited from a register of people who had given prior consent to be contacted about research studies and had been screened on a self-report measure of MDD: the World Health Organization’s Composite International Diagnostic-Short Form [12]. Table 2 displays the characteristics of this sample in terms of their gender and age, as well as the time postdiagnosis for each health condition.

Table 2. Sample characteristics across the three health conditions.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Major depressive disorder (n=8)</th>
<th>Epilepsy (n=7)</th>
<th>Multiple sclerosis (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female), n (%)</td>
<td>5 (63)</td>
<td>5 (71)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>51.9 (9.4)</td>
<td>44.4 (15.8)</td>
<td>43.4 (9.5)</td>
</tr>
<tr>
<td>Time postdiagnosis (years), mean (SD)</td>
<td>8.3 (10.3)</td>
<td>19.1 (16.2)</td>
<td>2.9 (1.6)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>5 (63)</td>
<td>6 (86)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Black</td>
<td>2 (25)</td>
<td>N/A b</td>
<td>N/A</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (13)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
<td>1 (14)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Theme-checking group follow-up, n (%)</td>
<td>6 (75)</td>
<td>5 (71)</td>
<td>5 (56)</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

**Focus Group Procedure**

A local research ethics committee (REC) approved these procedures (REC reference No. 16/LO/1513). All participants were screened for their eligibility to take part and, if eligible, were invited to a focus group session, for which travel expenses were covered. In this session, they first completed a consent form and a demographics questionnaire. We conducted separate focus groups for people with a diagnosis of MDD, MS, and epilepsy. For each, the main discussion was semi-structured using a prespecified topic guide (available on request). The discussion was designed based on the existing literature and through consultation with health care professionals and service users to elicit ideas about what was important to people in terms of their physical and mental health and well-being (eg, whether measuring relapse was important). In the topic guide, we referenced long-term conditions but also focused separately on symptoms of MDD, symptoms of MS, and seizure occurrence, tailoring this to the ones most relevant to the group. The open-discussion format allowed people to share a breadth of experiences, including what was important to their health and well-being, as well as suggestions for important areas to measure using RMT. Each group’s main discussion lasted 60-120 minutes and was combined with a conversation about potential barriers and facilitators to engagement, the content of which has been published elsewhere [13-15]. We invited all participants to comment on the themes extracted from the main discussion in a second focus group. This member-checking process allowed us to validate the themes that had been extracted from the main discussion. The results from this second session were combined with the first session to add further depth and to clarify the points raised.

**Data Analysis**

Focus group discussions were audio recorded and transcribed verbatim. For each health condition, analyses were conducted by two researchers working independently using the software package NVivo 11 (QSR International) [16]. Themes emerging in coding were resolved as a pair, and a joint decision was made about the allocation of a code to each quotation.

**Results**

**Overview**

A total of 24 participants, ranging from 28 to 65 years of age, took part in three focus groups; 67% (16/24) of participants were women. Of the 24 participants, 16 (67%) returned for a further member-checking session to verify the findings. This meant that six focus groups were run in total. There was a similar distribution of men and women across the focus groups. However, participants with a history of MDD were, on average, slightly older. The time spent living with the chronic health condition varied; the people with MS had, on average, been living with their condition for the shortest amount of time.

The focus groups identified several factors important to health and well-being across the three health conditions. We have divided these results into the measurement priorities important for each clinical group separately. For MDD and MS, the discussions centered around the importance of detecting signs of relapse or deterioration in health; for epilepsy, the focus was on the detection of seizures. For all groups, there was a consideration of how RMT may support well-being as well as symptoms of illness and contextual factors.

**Major Depressive Disorder**

Participants were asked what was important to their health and well-being and what factors may be important to measure using RMT. They reported a plethora of possible symptoms commonly associated with relapse in MDD, including negative thoughts (ie, about being dissatisfied with themselves, unsupported, and
burnt out); poor sleep; changes in appetite (ie, for some, this included experience of eating disorders); withdrawal from activities, including social activities and self-care; and anxiety, including fear of relapse.

_I was thinking probably when I don’t sleep well … that’s a sign. You can get these tracker things now and I was thinking getting one myself, that’s supposed to track your sleep. I thought maybe something as simple as that might actually be helpful._ [MDD participant #8, regarding poor sleep as a sign of relapse]

In addition to relapse, some participants valued a focus on remission or maintenance of wellness. For measurable signs of wellness, participants had several suggestions, including being more active, such as participating in more social and other leisure activities (ie, moderate physical activity) and engaging in employment; eating well; feeling in control and actively coping with situations; feeling good about oneself; and experiencing a sense of achievement.

_I like recording what keeps me well, not what makes me ill. I’d much prefer contemplating to think more positively. To think, “oh these things work.” I like to keep focused on the positive side._ [MDD participant #4, regarding measuring wellness]

Contextual factors that included life events, such as bereavement, problems with employment, and financial difficulties, were seen to be important to monitor. Additional physical health problems and availability of support in the context of barriers, such as social isolation, were mentioned as potentially stress-inducing contextual factors. One person mentioned the importance of tracking information that might be useful for medication management.

_I could see that if um the tracking information would be useful for my doctor, to help with trying to find the right medication._ [MDD participant #6, regarding the importance of tracking medication use]

**Multiple Sclerosis**

Participants with a diagnosis of MS also endorsed using RMT to measure and predict relapse but mostly in the context of a diagnosis of relapsing-remitting MS. For participants with a diagnosis of progressive forms of MS, relapse was less important because this did not reflect their experience of living with their condition.

_The other thing I’d find useful would be to be able to sort of track how much worse I’m getting, it’s very hard to know, because it’s very gradual in a way, the deterioration I’m getting._ [MS participant #8, regarding measuring deterioration]

This suggests that a focus on change in severity of symptoms would still be of importance to measure when using RMT. Deterioration in mobility and gait were key symptoms highlighted. However, participants emphasized the importance of measuring additional symptoms, such as vision (ie, for some, optic neuritis was an early symptom of MS relapse), fatigue, and social functioning. Mental health was also thought to be important to measure. Participants highlighted specific times that may be associated with greater distress, including the time before their diagnosis, and periods of relief afterward. These key moments in the trajectory of people’s illness may be particularly important targets for remote measurement and intervention.

In addition to symptoms of illness, some participants spoke of the value in measuring signs of wellness, for instance, eating well and being active. Individual contextual factors such as outlook or attitude modified their experience, with active attempts to cope being potentially protective for well-being.

_If there’s something that monitors everything that you’ve eaten that day and what you’ve been doing that day and then it’s like, “okay that’s been a good day,” then you’ll have that information to think, “well maybe I’ll do more of that to try and increase the amount of good days.”_ [MS participant #9, regarding measuring wellness]

**Epilepsy**

Participants with epilepsy saw the potential importance of RMT in its ability to measure the frequency of seizures, as well as presesizure symptoms or predictors. The unpredictable nature of seizure occurrence was discussed among participants, including the potential value for technology to provide more control.

_I get warnings before my seizures but they’re not very long, so if I can predict it even before that, it might change the way I plan my day._ [Epilepsy participant #6, regarding value of predicting seizures]

Perhaps due to the uncertainty surrounding predictors of seizure, different participants raised different parameters of importance. Those most frequently mentioned included change in emotions, including anger, anxiety, and more positive emotions such as excitement, as well as altered sleep, including sleep deprivation and irregular sleep patterns. Physiological signals, such as heart rate and brain activity (eg, electroencephalography), were mentioned to help detect seizures.

It is important to note that some participants felt that a singular focus on seizures may be problematic. Participants spoke of epilepsy having an impact on their life in a more holistic way. Contextual factors such as effects on working life may be just as important to track as seizure frequency. These contextual factors were framed in terms of the losses that people with epilepsy experience as a result of their health condition (eg, loss of employment).

_It’s actually the 23 hours of every day when you’re not having a fit, that’s the time that the epilepsy has the biggest effect._ [Epilepsy participant #4, regarding importance of context]

_I don’t want that constant reminder when I’m having a good day._ [Epilepsy participant #6, regarding importance of the ability to forget diagnosis on well days]

Despite the importance of a holistic approach, the group did not think that focusing on signs of wellness would always be of
help. One person stated that it might be annoying to be constantly reminded that they had a diagnosis of epilepsy on days when they felt well. This linked to a discussion that acceptance of their own health condition was hard and potentially influenced by a felt sense of stigma. The psychosocial impact of epilepsy may be important to track.

**Comparison Across Health Conditions**

From Figure 1, it is apparent that the views of people with MDD, epilepsy, and MS differed. Each group highlighted unique measurements of importance, relevant to their specific needs. Although there was agreement that remote measurement could be useful for tracking symptoms of illness, some symptoms were specific to the individual groups: for MDD this included negative thoughts; for MS it was reduced mobility and poor vision; and for epilepsy it was change in physiological parameters, such as heart rate and activity in the brain. That said, some symptoms were shared across the groups, including poor sleep (MDD and epilepsy), reduced social functioning (MDD and MS), as well as diet and anxiety (MDD, epilepsy, and MS). Measuring signs of wellness were mentioned more by people with MDD than by people with MS and epilepsy. However, there was some overlap between MDD and MS, with increased activity being important to both. Overlap also emerged when considering contextual factors, such as life events, and availability of support (MDD and epilepsy), as well as ways of coping (epilepsy and MS).

**Figure 1.** The unique and overlapping outcomes of importance for three chronic health conditions: major depressive disorder (ie, depression), epilepsy, and multiple sclerosis. Grey areas outside of the overlapping sections represent contextual factors either shared or uniquely mentioned by members of the focus groups.

### Discussion

#### Principal Findings

When participants in this consultation exercise were asked what they thought would be a successful measurement for the implementation of RMT, they endorsed the idea of detecting and predicting relapse (for MDD and relapsing-remitting MS) or negative change in health state (ie, deterioration for progressive forms of MS and seizure occurrence for epilepsy). Symptoms of relapse or negative change in health as described in the focus groups have been well documented; they form the basis of clinical assessment interviews and self-report tools that have been validated to measure severity of MDD (eg, the nine-item Patient Health Questionnaire) [17], MS (eg, the UK Functional Independence Measure and Functional Assessment Measure) [18], and epilepsy (eg, the Liverpool Seizure Severity Scale) [19]. These are very clearly measurements of interest for studies using RMT. If symptoms can be identified early, timely interventions may be offered, before these symptoms become more severe.

A Holistic and Context-Specific Approach

It is important to view the conditions MDD, MS, and epilepsy both separately and holistically, meaning that we choose end points that can help us to understand people as unique individuals experiencing complex health conditions and environments. People with MDD did not only want to be monitored for symptoms of MDD, but also anxiety. In addition, they wanted to measure their physical health. This is in line with existing research on the importance of physical health as a risk factor for MDD [20-22]. For people with MS and epilepsy, the combination of measuring mental health as well as physical health emerged too. MDD and anxiety are prevalent disorders among people with both MS [23] and epilepsy [24,25], and may contribute to early signs of relapse or deterioration in the health state. Using RMT to actively measure fluctuations in mood disorder and anxiety in real time may help to gather more reliable findings. RMTs are uniquely positioned to be able to address problems with recall bias introduced when there is a delay in self-reporting of experiences.

For MDD and MS, there were discussions about maintaining wellness and what this looked like, most commonly, in terms of increased activity, positive social functioning, and access to...
support. There may be an argument for including real-time measures of well-being (eg, the Warwick-Edinburgh Mental Well-being Scale) [26] and quality of life (eg, the EuroQol five-dimension questionnaire) [27] for RMT studies conducted for these groups. Passive measures of functioning gained through an analysis of mobile phone usage and wearable devices (eg, call logs and step counts) may also be of value. For people with epilepsy, there was little focus on maintaining wellness; people spoke about their illness being out of their control with unpredictable triggers in terms of how they were living their life. Difficulties establishing triggers for seizures has been a well-documented finding within the previous literature [28]. For people with epilepsy, being able to receive a warning of their seizure early was of most importance to them. This highlights a difference between the needs of people with epilepsy compared to the two other chronic health conditions.

**Strengths and Limitations**

The strengths of this study include the opportunity for an open and in-depth discussion with people who have first-hand experience of living with one of three chronic health conditions. This enabled a rich exploration of the health measurements of importance and allowed us to identify similarities and differences between the groups. The employed member-checking methods allowed validation of the results generated from the main discussion. Given the qualitative approach, we are limited in our ability to quantify the numbers of people wanting to measure specific outcomes or to run any statistical analyses to explore the significance of group differences, including factors such as diagnosis, age, ethnicity, and other characteristics not quantified, like the previous use of mHealth resources and income. This work has generated ideas that will inform the design of RMT studies. These RMT studies will test the relationships between the measurements of interest, including those identified in these focus groups.

**Conclusions**

In this consultation exercise, we identified measurements of importance when using RMT for three chronic health conditions: MDD, MS, and epilepsy. This is a unique study that puts patients’ views at the forefront of the design of a clinical study employing novel digital resources. We draw the following conclusions. First, in all cases, measuring symptom severity is key: people want to know when their health is getting worse. Second, symptom severity needs to be placed in context. When monitoring someone with a mental health condition such as MDD, social and physical health outcomes should also be considered, and vice versa for physical health conditions such as MS and epilepsy. A holistic approach that considers situational and attitudinal factors (eg, employment, social status, acceptance of health condition, eating patterns, and ways of coping) will enable a more complete picture of how unwell a person is feeling. For some people with MDD and MS, factors that maintain well-being are just as important as factors that contribute to relapse or deterioration in health status.

### Acknowledgments

This paper was written as part of the development of useful mHealth and RMT systems in the Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) project. The RADAR-CNS project has received funding from the Innovative Medicines Initiative (IMI) 2 Joint Undertaking (grant agreement No. 115902). This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA) [29] with one pharmaceutical company supporting the preparation of this manuscript as part of this pre-competitive public and private partnership. This communication reflects the views of the RADAR-CNS Consortium [30], and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. This paper also represents independent research partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. Professor Til Wykes (TW) would specifically like to acknowledge support from the NIHR for their Senior Investigator Awards.

### Authors’ Contributions

TW, MH, and VN conceived of the project and oversaw how the study was conducted. SS conducted the research with the assistance of FM, BG, and HC and wrote the first draft of the manuscript. Further contributions to the writing of this draft were made by JN, AP, JF, and PG. All authors reviewed and commented on the final draft.

### Conflicts of Interest

None declared.

### References


https://mhealth.jmir.org/2020/6/e15086
Promoting Antenatal Care Attendance Through a Text Messaging Intervention in Samoa: Quasi-Experimental Study

Jessica L Watterson1,2, MPH, PhD; Diego Castaneda3, MPH, DrPH; Caricia Catalani2, MPH, DrPH

1School of Public Health, University of California, Berkeley, Berkeley, CA, United States
2Design for Health, IDEO, San Francisco, CA, United States
3Facebook Research, Menlo Park, CA, United States

Corresponding Author:
Jessica L Watterson, MPH, PhD
School of Public Health
University of California, Berkeley
2121 Berkeley Way
Berkeley, CA, 94720
United States
Phone: 1 3602014410
Email: jessica.watterson@gmail.com

Abstract

Background: Antenatal care (ANC) has the potential to improve maternal health, but it remains underutilized and unevenly implemented in many low- and middle-income countries. Increasingly, text messaging programs for pregnant women show evidence that they can improve the utilization of ANC during pregnancy; however, gaps remain regarding how implementation affects outcomes.

Objective: This study aimed to assess facilitators and barriers to implementation of an SMS text messaging intervention for pregnant women in Samoa and to assess its impact on ANC attendance.

Methods: This study took place in Upolu, Samoa, from March to August 2014 and employed a quasi-experimental design. Half (n=3) of the public antenatal clinics on the island offered adult pregnant women the SMS text messaging intervention, with 552 women registering for the messages. At the comparison clinics (n=3), 255 women registered and received usual care. The intervention consisted of unidirectional text messages containing health tips and appointment reminders. The outcome of interest was the number of attended antenatal visits. Data analysis included a comparison of women’s baseline characteristics between the two groups, followed by the use of negative binomial regressions to test for associations between participation in the intervention and increased ANC attendance, controlling for individual characteristics and accounting for the clustering of women within clinics.

Results: The comparison of ANC attendance rates found that women receiving the SMS text messaging intervention attended 15% fewer ANC visits than the comparison group (P=.004), controlling for individual characteristics and clustering. Data analysis of the implementation process suggests that barriers to successful implementation include women registering very late in pregnancy, sharing their phone with others, and inconsistent explanation of the intervention to women.

Conclusions: These results suggest that unidirectional text messages do not encourage, and might even discourage, ANC attendance in Samoa. Interpreted with other evidence in the literature, these results suggest that SMS text messaging interventions are more effective when they facilitate better communication between patients and health workers. This study is an important contribution to our understanding of when SMS text messaging interventions are and are not effective in improving maternal health care utilization.

(JMIR Mhealth Uhealth 2020;8(6):e15890) doi:10.2196/15890

KEYWORDS
mHealth; antenatal care; maternal health; text messages
Introduction

Background
As an independent state, Samoa has achieved high performance on indicators of maternal health, including relatively high rates of deliveries in medical facilities (82%) and high rates of women receiving antenatal care (ANC, 93%) [1]. However, only 73% of women receive four or more antenatal visits, as recommended by the World Health Organization, and only 12% of women register for care in the first trimester [1]. To improve maternal health, rates of early, regular ANC attendance should be improved. Antenatal interventions, particularly those focused on chronic conditions (eg, anemia, infections, and hypertensive disorders), have the potential to detect, treat, or prevent conditions that could otherwise lead to maternal mortality or morbidity [2]. Samoa’s maternal mortality ratio was estimated at 100 maternal deaths per 100,000 live births in 2010 [3]. The Ministry of Health’s Antenatal Care Survey in 2012 found that many mothers did not think they needed to attend ANC because they felt their baby was safe and in good health (23%; Samoa Ministry of Health, unpublished data, 2012). These results indicate that the importance of ANC must be emphasized to pregnant women to ensure they attend ANC, even if they feel healthy.

Text message reminders and education interventions for pregnant women have been implemented widely around the world, but relatively few have been systematically evaluated to determine their effects on maternal care-seeking behavior or health outcomes. Among the studies that have examined outcome measures, there is some evidence that SMS text messaging programs can improve health care utilization, knowledge, and satisfaction with care. For example, Lund et al [4] conducted a pragmatic randomized controlled trial (RCT) in Zanzibar and found that women receiving unidirectional text messages and free mobile phone credit to communicate with their health provider had double the odds of attending four or more antenatal visits, relative to the control group. Similarly, studies in Malawi and Iraq found increased ANC attendance among women who received a unidirectional text or voice messaging intervention and access to hotlines or phone numbers to call with questions [5,6]. A recent meta-analysis of seven RCTs in low- and middle-income countries found evidence that text messages for pregnant women significantly increased ANC attendance by 174% [7]. Other studies have also found SMS text messaging interventions to increase mothers’ knowledge, preparedness, feelings of empowerment, and satisfaction with ANC [7-10].

Samoa provides a promising context in which to study text messages for maternal health because an estimated 90% of the population of Samoa had access to a mobile phone in 2013 [11], and nearly 99% of the adult population is literate [3]. In addition, free ANC is provided at public health facilities across the country. Although a handful of studies have found evidence for the effectiveness of SMS text messaging programs at increasing ANC attendance, more evidence is needed to understand in what environments these programs can produce results for women’s health [12]. Previous studies have examined the outcomes of these programs in countries in Africa and Asia with different cultures, religions, literacy rates, incomes, and health care systems—all factors that could contribute to or detract from the effectiveness of a pregnancy SMS text messaging intervention. Therefore, this study explores whether this intervention can be effective in the Samoan context, contributing to a more nuanced understanding of how the setting and implementation factors might affect the outcomes of a pregnancy SMS text messaging program.

Hypotheses
On the basis of earlier findings that SMS text messaging interventions have been successful at improving ANC attendance in other developing countries, we hypothesized the following:

1. Pregnant women receiving the SMS text messaging intervention will attend a higher number of follow-up antenatal visits than women not receiving them, controlling for other individual characteristics;
2. The SMS text messaging intervention will have a greater effect on younger pregnant women’s ANC attendance compared with older women, controlling for other individual characteristics.

Evidence from around the world indicates that younger people tend to have higher rates of mobile phone ownership and higher technological literacy [13,14], suggesting that the effect of an SMS text messaging intervention could be even greater for younger women. In addition, these women are more likely to be first-time mothers and to be interested in additional supportive information, such as that provided by the SMS text messaging program.

Methods

Study Design
This study was conducted from March to September 2014 in Samoa. The study took place on the island of Upolu, the most populated island and home to the capital city, Apia. The National Health Service runs 6 health centers on the island (1 urban and 5 rural), all offering free antenatal services to pregnant women. Ethics approval for this study was obtained from the National Health Research Committee of Samoa on February 6, 2014. Per the approved protocol, participants receiving the intervention provided verbal consent to participate in the study to the clinic midwives. Analysis of the deidentified dataset was deemed to be not human subjects research by the University of California Berkeley Office for the Protection of Human Subjects on September 7, 2017.

This study employed a cluster-randomized quasi-experimental study design, in which half of the health centers (n=3) were randomly selected to offer the SMS text messaging intervention to pregnant women presenting for their first antenatal visit, and the other half of the clinics (n=3) were randomly selected to offer the usual care only. Random selection was performed by assigning each clinic a number from 1 to 6 and then using a Web-based random number generator to select 3 of the numbers randomly to identify intervention clinics. Figure 1 illustrates the locations of these clinics on Upolu Island. Nurse-midwives offered the SMS text messaging intervention to pregnant women...
(n=728) who registered at an intervention clinic. Pregnant women who registered at comparison clinics during the study period were enrolled in the comparison group (n=251).

The study included a total of 979 women, all of whom registered at 1 of the 6 public antenatal clinics in the study period. The only pregnant women not eligible for inclusion in the study during this period were those who did not attend ANC in a clinic (eg, those visiting a traditional birth attendant, estimated at 3% of pregnant women [1]), or those who visited a private health care provider. This is a relatively small percentage of the population, on the basis of the significantly higher cost and limited reach of most private facilities (most are located in the capital city, Apia).

Figure 1. Map of National Health Service clinics in Upolu, Samoa. Circles indicate a comparison clinic and triangles indicate an intervention clinic.

Women in the intervention group received 2 educational messages per week, with content adapted to their gestation (eg, if a woman was 20 weeks pregnant at registration, the first educational messages she received were adapted for 20 weeks of pregnancy, then 21 weeks the following week, and so on). All women in the intervention group received the same educational messages at the same gestational age (ie, the messages were not tailored to individuals). The text messages were adapted for the local context and translated to Samoan from the free library developed by the Mobile Alliance for Maternal Action on the basis of the Lancet Maternal and Neonatal Survival Series. Adaptations included removing content about malaria (malaria is not endemic in Samoa) and ensuring fruits and vegetables that were referenced were familiar and locally available. Women in the intervention group also received a text message appointment reminder the day before their scheduled appointment. Finally, a reminder message was sent to women who were over 4 weeks overdue for an appointment.

This study examined the effect of text message education and reminders on ANC attendance, which was measured by the number of follow-up ANC visits attended. This outcome was selected on the basis of earlier research that found SMS text messaging programs showed promise for improving ANC attendance in other settings [4-6]. Data were collected from medical records and ANC registration books in antenatal clinics. Although women were of different gestations at registration, and therefore had different recommended antenatal schedules, gestational age was controlled for in the multivariate analyses. All available demographic information was also collected from medical records for each woman, including her age, marital status, parity, and whether she or her partner was employed outside the home and her home village. These demographic details are comparable with those included in similar studies and are thought to be potential confounders for ANC attendance, which is why they were included in the analysis. Details on the variables collected for both the intervention and control groups are outlined in Table 1. A survey for implementation feedback was also conducted about 1 month after beginning the program with the implementing midwives (n=7), and the researcher maintained detailed implementation notes and records.
Table 1. Description of variables and data sources.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Age in years at time of ANC registration</td>
<td>Medical record or ANC registration book</td>
</tr>
<tr>
<td>Parity (including current pregnancy)</td>
<td>Total number of pregnancies, including current pregnancy</td>
<td>Medical record</td>
</tr>
<tr>
<td>Distance from home village to registration clinic (km)</td>
<td>Distance from home village to the clinic where the woman registered for ANC in kilometers (km)</td>
<td>Home village recorded from medical record or ANC registration book, then distance from the registration clinic in km was estimated using Google Maps</td>
</tr>
<tr>
<td>Married/in partnership</td>
<td>Marital status recorded as married or stable union</td>
<td>Medical record</td>
</tr>
<tr>
<td>Employed and/or partner employed</td>
<td>Occupation of the pregnant woman and/or husband/partner was recorded, then categorized as being at home or outside the home. If 1+ person worked outside the home, they were categorized as employed</td>
<td>Medical record</td>
</tr>
<tr>
<td>Gestation at registration (weeks)</td>
<td>Number of weeks pregnant at the time of registration for ANC</td>
<td>Medical record or ANC registration book</td>
</tr>
<tr>
<td>Number of follow-up antenatal visits attended</td>
<td>Number of visits attended after the first registration visit; dates of subsequent visits were recorded, then counted</td>
<td>Medical record or ANC registration book</td>
</tr>
<tr>
<td>Intervention group</td>
<td>Enrolled in the intervention group if women were pregnant, over 18 years of age, presented to an intervention clinic for ANC registration, and agreed to participate</td>
<td>Sign-up sheet or registration book from midwives in clinic</td>
</tr>
</tbody>
</table>

ANC: antenatal care.

The required sample size was estimated first without accounting for clustering, as in similar studies [4]. To detect a difference of one follow-up antenatal visit between the intervention and control groups with alpha=.05 and beta=.10, a sample size of 262 is needed (n=131 per group). This estimate was on the basis of a conservative approximation of the effect size and standard deviation found in a study by Alhaidari et al [6], which also examined the effect of an SMS text messaging intervention on the number of ANC visits attended. Although this estimate did not take clustering into account, it was known that the final sample size would be significantly larger given Samoa’s birth rate, the population of Upolu, the length of the study, and the high percentage of pregnant women who attend at least one ANC visit with a health care provider [1].

Missing Data

Problems with locating complete paper medical records led to one or more missing demographic variables for 214 participants. Varied filing systems, large volumes of patients seen each day, and many common names led to difficulty locating patients’ records, both for the researcher and for clinic staff. The distribution of this missing data is outlined in Table 2 in the Results section. The missing data were relatively evenly distributed across both intervention and comparison groups, reducing concerns about bias. The main analyses used listwise deletion of these observations with missing values (106 from the per-protocol intervention group and 108 from the per-protocol comparison group). As a sensitivity analysis, multiple imputations by chained equations was used to generate 20 datasets with 975 complete observations each, which were then combined for analysis using Rubin combination rules [15]. Data were imputed for the variables age at registration, parity, marital status, and employment status. Data were not imputed for the 4 observations missing the distance from their home village to the registration clinic because of the high correlation of this variable with other variables (the model did not achieve convergence). The number of imputed datasets was determined using the proportion of missing data and acceptable power falloff [16]. The sensitivity analysis then proceeded with the same models as the main analysis (described below in the Data Analysis section), and the results were compared.

Data Analysis

Statistical analyses were conducted using the Stata/SE 13.0 software (Stata Corp LP). Basic descriptive statistics were calculated for all variables and separately for the intervention and comparison groups. This included means, medians, and standard deviations for all continuous variables and frequencies, proportions, and 95% CIs for all categorical variables. Descriptive statistics for both groups were compared using t tests for continuous variables and chi-square tests for categorical variables.

The intervention and comparison groups were categorized using both the intention-to-treat and per-protocol principles. In the intention-to-treat analysis, all women registering for ANC at an intervention clinic were treated as receiving the intervention, regardless of whether they signed up to receive the text messages or not. In the per-protocol analysis, the women who did not actually receive any text messages were considered part of the comparison group, regardless of at which clinic they registered. To study the significance of differences in the number of antenatal visits that were attended between the two groups, negative binomial regressions were estimated, controlling for patient demographics and accounting for clustering within clinics using a clustered sandwich estimator to produce robust standard errors. Next, the same model was run with an interaction term for young women (defined as under the age of 25) and being in the intervention group.
Implementation survey data were analyzed by calculating basic descriptive statistics for quantitative questions. Open-ended responses to survey questions and implementation notes were carefully reviewed to identify common themes.

**Results**

**Descriptive Statistics**

Figure 2 outlines the results of the study registration. A total of 728 women registered for ANC at 1 of the 3 intervention clinics during the study period. Of these women who were offered the SMS text messaging intervention, 75.8% (552/728) signed up. The majority of women who registered at an intervention clinic but did not receive the text messages registered very late in pregnancy (ie, within 2 weeks of their due date), or their phone number was not recorded so messages could not be sent (n=49). A total of 127 women elected not to receive the text messages, and 18 of those women did not have a mobile phone. A total of 251 women were registered at a comparison clinic during the study period.

Challenges locating complete paper records led to one or more missing demographic variables for 214 participants. The distribution of this missing data is outlined in Table 2. These observations with missing values were excluded from subsequent analyses.

Figure 2. Antenatal care and text message (SMS) intervention registration results (N=979).

Table 2. Distribution of missing observations across groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intention-to-treat, n (%)</th>
<th>Per-protocol, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention (n=728)</td>
<td>Comparison (n=251)</td>
</tr>
<tr>
<td>Age</td>
<td>128 (17.6)</td>
<td>45 (17.9)</td>
</tr>
<tr>
<td>Parity (including current pregnancy)</td>
<td>127 (17.4)</td>
<td>47 (18.7)</td>
</tr>
<tr>
<td>Distance from home village to registration clinic (km)</td>
<td>4 (0.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Married/in partnership</td>
<td>94 (12.9)</td>
<td>42 (16.7)</td>
</tr>
<tr>
<td>Employed and/or partner employed</td>
<td>147 (20.2)</td>
<td>61 (24.3)</td>
</tr>
<tr>
<td>Missing any of the above variables</td>
<td>151 (20.7)</td>
<td>63 (25.1)</td>
</tr>
</tbody>
</table>

Descriptive statistics for both intervention and comparison groups with complete data according to both intention-to-treat and per-protocol categorization are outlined in Table 3. The size of the intervention group was larger than that of the comparison group because of the inclusion of the antenatal clinic in the main hospital as an intervention site (Tupua Tamasese Meaole Hospital). This clinic saw the highest number of women registering for ANC, which resulted in the larger intervention group. The demographic characteristics of women in the intervention and comparison groups were similar at baseline, with two exceptions. The proportion of women and/or their partners who were employed outside the home was significantly higher in both intervention groups, regardless of whether they were categorized according to per-protocol or intention-to-treat. Similarly, the mean distance traveled by women from their home village to the clinic they registered at was higher among the intervention groups than in the comparison groups, although these distances varied widely (from 0 to 117 km), and thus have high standard deviations. This was also likely because of the inclusion of the main hospital as an intervention site, as women are more likely to have traveled from a rural area to the capital city to attend their appointment there.

Women in the per-protocol intervention group received a mean of 25.6 messages throughout the intervention (SE 0.47), and women in the intention-to-treat intervention group received a lower mean of 19.3 messages (SE 0.54) because of the fact that 176 of these women received no messages, despite being randomized to receive them (results not shown in the table).
Table 3. Baseline characteristics of intervention and comparison groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intention-to-treat</th>
<th>Per-protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention (n=577)</td>
<td>Comparison (n=188)</td>
</tr>
<tr>
<td>Continuous variables, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>26.7 (6.4)</td>
<td>27.1 (6.5)</td>
</tr>
<tr>
<td>Parity (including current pregnancy)</td>
<td>3.2 (2.0)</td>
<td>3.3 (2.0)</td>
</tr>
<tr>
<td>Distance from home village to registration clinic (km)</td>
<td>11.9 (13.1)</td>
<td>6.6 (7.2)</td>
</tr>
<tr>
<td>Gestation at registration (weeks)</td>
<td>27.2 (6.7)</td>
<td>26.5 (6.0)</td>
</tr>
<tr>
<td>Number of follow-up antenatal visits attended</td>
<td>2.2 (1.9)</td>
<td>2.6 (1.7)</td>
</tr>
<tr>
<td>Categorical variables, n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/in partnership</td>
<td>519 (89.9)</td>
<td>171 (91.0)</td>
</tr>
<tr>
<td>Employed and/or partner employed</td>
<td>405 (70.2)</td>
<td>89 (47.1)</td>
</tr>
</tbody>
</table>

*Excluding missing data.

Comparison of Antenatal Care Visits Attended for Intervention and Comparison Groups

Using the intention-to-treat principle, women registering at intervention clinics attended, on average, only 2.2 follow-up visits, as compared with 2.6 in the comparison group (*P*=.01). Similarly, in the per-protocol analysis, women receiving the intervention attended only 2.1 follow-up visits on average, compared with 2.5 visits in the comparison group (*P*<.001). These unadjusted comparisons are presented in Table 3.

Contrary to hypothesis 1, the negative binomial regression analyses (Table 4) showed that women in the intervention group attended 13% (intention-to-treat) to 15% (per-protocol) fewer follow-up ANC visits than women in the comparison group, controlling for all covariates. The interaction term between younger women (defined as under 25 years old) and receiving the intervention in the subsequent regression model was not significant (*P*=.30), suggesting that the effect of the intervention on ANC attendance was similar across age groups (results not shown in table). Therefore, support was not found for hypothesis 2.

As a sensitivity analysis, the multivariate regressions were run again with the 20 multiply imputed datasets (n=975). The estimated effect of receiving the intervention on the number of follow-up ANC visits attended was slightly smaller (ie, the incidence rate ratio [IRR] was closer to 1.0) and no longer statistically significant in these regression results (IRR=0.88, *P*=.06, per-protocol).

Table 4. Comparison of visits attended between intervention and comparison groups, controlling for demographic characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intention-to-treat</th>
<th>Per-protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IRR</td>
<td>Robust SE</td>
</tr>
<tr>
<td>Intervention group</td>
<td>0.87</td>
<td>0.06</td>
</tr>
<tr>
<td>Age at registration</td>
<td>1.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Married/in partnership</td>
<td>0.95</td>
<td>0.10</td>
</tr>
<tr>
<td>Parity</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td>Employed and/or partner employed</td>
<td>0.91</td>
<td>0.06</td>
</tr>
<tr>
<td>Distance from home village to registration clinic (km)</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Gestation at registration (weeks)</td>
<td>0.99</td>
<td>0.00</td>
</tr>
<tr>
<td>Constant</td>
<td>3.36</td>
<td>0.58</td>
</tr>
</tbody>
</table>

*aIRR: incidence rate ratio.

Midwife Survey and Qualitative Results

The survey of implementing midwives (Table 5) indicated that they found the program to be useful (mean score of 4.0 out of 5). The average rating of how interested they thought their patients were in receiving the messages was lower (mean score of 3.1 out of 5) compared with other questions. In addition, the midwives felt that registering pregnant women for the messages (which involved recording the woman’s name, phone number, and gestation on a form) was fairly difficult (mean score of 4.29 out of 5, where 5 is difficult).
Analyses of qualitative data from implementing midwives and implementation notes identified facilitators and barriers to a successful implementation of the SMS text messaging program. A key barrier was difficulty with consistently offering and explaining interventions to women at intervention clinics. Despite the implementing midwives participating in training at the program’s start and regular visits from the researcher to discuss the program and collect data, evidence suggests that some pregnant women might not have received a clear explanation of the program, or might not have been offered the program even if they registered for ANC at an intervention clinic. One midwife wrote, “I sometimes forget to fill in forms but will improve as it becomes part of daily routine.” This quote highlights that implementation of the intervention did not fit into the midwives’ existing workflow, which might have contributed to inconsistent registration, and could explain why the midwives rated registering pregnant women for the program as fairly difficult. In addition, the researcher received responses to some of the text messages asking who had sent the message. This could suggest potential issues such as (1) someone else was using the mobile phone, as phone sharing is a common practice among friends and families in Samoa, or (2) the woman had not understood or had forgotten that she signed up for the messages at the clinic.

One of the key facilitators identified was offering the option for women to enroll in the message by paper during their ANC visit, rather than requiring them to send an SMS text message to enroll. Many mobile messaging platforms enroll participants by having them send a short code to a phone number. However, this can cost the participant’s mobile phone credit, to ask questions. In fact, this study found some evidence that women receiving the unidirectional messages attended fewer follow-up ANC visits than did women not receiving the messages, controlling for individual characteristics and clustering within clinics. One potential explanation for this finding could be that the messages led participants to feel more connected to the clinic, or that they had sufficient information, reducing their motivation to attend an in-person check-up (ie, there was a substitution effect, whereby patients substituted information received by text message for more time-intensive ANC). Further study is needed to understand the components of SMS text messaging programs that encourage (or discourage) ANC attendance and whether adjustments to the implementation (eg, features, content, and scheduling) could impact the effectiveness of the intervention in improving attendance.

Comparison With Prior Work
To date, only a handful of studies have examined outcome measures for SMS text messaging interventions for maternal health and found positive results, and each of these studies included some features beyond what our intervention offered [12]. For example, a study in Sierra Leone found an increase of 11.3% in attendance at the fourth antenatal visit after implementation of a bidirectional SMS text messaging intervention that allowed pregnant women to communicate with health care workers [17]. Similarly, a study in Malawi found an increase in antenatal attendance after implementing a case management hotline and unidirectional text and voice messaging [5]. A recent literature review of studies using SMS text messaging for maternal and infant health found evidence that bidirectional messaging might be more effective [18]. Taken together, previous evidence and our study indicate that interventions may need to increase bidirectional interaction with pregnant women and move beyond unidirectional reminders and health tips. Enhancing patient engagement may enable text-based interventions to have a greater impact on patient care-seeking behavior. On the basis of our implementation findings, another idea to explore in future research is whether the participation of women’s partners, family members, or friends could improve the program’s outcomes. A recent meta-analysis of male involvement during pregnancy found evidence of improved utilization of maternal health services [19], lending further support to the idea that women’s social networks could support them to attend more antenatal visits if involved in the SMS text messaging program.

Discussion

Principal Findings

Despite some previous evidence for the effectiveness of SMS text messaging interventions for increasing attendance to antenatal visits, our results indicate that they may not necessarily be effective at improving health-seeking behavior when implemented in isolation of other interventions, such as hotlines or phone credit, to ask questions. In fact, this study found some evidence that women receiving the unidirectional messages attended fewer follow-up ANC visits than did women not receiving the messages, controlling for individual characteristics and clustering within clinics. One potential explanation for this finding could be that the messages led participants to feel more connected to the clinic, or that they had sufficient information, reducing their motivation to attend an in-person check-up (ie, there was a substitution effect, whereby patients substituted information received by text message for more time-intensive ANC). Further study is needed to understand the components of SMS text messaging programs that encourage (or discourage) ANC attendance and whether adjustments to the implementation (eg, features, content, and scheduling) could impact the effectiveness of the intervention in improving attendance.

Comparison With Prior Work
To date, only a handful of studies have examined outcome measures for SMS text messaging interventions for maternal health and found positive results, and each of these studies included some features beyond what our intervention offered [12]. For example, a study in Sierra Leone found an increase of 11.3% in attendance at the fourth antenatal visit after implementation of a bidirectional SMS text messaging intervention that allowed pregnant women to communicate with health care workers [17]. Similarly, a study in Malawi found an increase in antenatal attendance after implementing a case management hotline and unidirectional text and voice messaging [5]. A recent literature review of studies using SMS text messaging for maternal and infant health found evidence that bidirectional messaging might be more effective [18]. Taken together, previous evidence and our study indicate that interventions may need to increase bidirectional interaction with pregnant women and move beyond unidirectional reminders and health tips. Enhancing patient engagement may enable text-based interventions to have a greater impact on patient care-seeking behavior. On the basis of our implementation findings, another idea to explore in future research is whether the participation of women’s partners, family members, or friends could improve the program’s outcomes. A recent meta-analysis of male involvement during pregnancy found evidence of improved utilization of maternal health services [19], lending further support to the idea that women’s social networks could support them to attend more antenatal visits if involved in the SMS text messaging program.
Limitations

Digital intervention research is still in its infancy, especially in the developing world, and as such, there were limitations that may have affected the effectiveness of this program. Future studies should take these into account to continue to improve our understanding of these interventions. First, cluster randomization (at the health center level) was preferred because it did not require health workers to keep track of which individuals received the intervention. In doing so, we were unable to randomize individual women to the intervention, which would have improved causal inference.

Second, data on the presence of pregnancy complications were not available. Pregnancy complications could have influenced the number of ANC appointments attended (eg, if a woman is experiencing complications, her midwife will encourage her to come for more frequent check-ups). However, women were assigned to the intervention or comparison group on the basis of which clinic they first presented to, regardless of later transfers to the main tertiary hospital in Apia (as would occur if a complication was identified). This suggests that complications might have also been evenly distributed across groups. However, we have no way to test this with our current dataset. Further, if any bias were introduced by the presence of more complications in one group, we would expect that women with more complications would have presented to the main tertiary hospital, which was an intervention site. Therefore, this would have likely biased our results such that the intervention group would have attended more visits than the comparison group (opposite to our findings).

Third, one intervention clinic was based at the main hospital in the capital city and thus was significantly larger than any of the other clinics. Women traveled from many rural parts of the island to receive ANC at this clinic, but we do not know if there are other systematic ways in which the women registering at this clinic are different from women registering elsewhere. We account for the potential of longer distances traveled and the clustering of women within clinics in the multivariate regression models in an attempt to address this issue.

Fourth, because of difficulties in locating paper medical records in many of the clinics, there was a significant amount of missing demographic data. Attendance data for these patients were still collected from registration books, so only demographic data were missing. The results of the sensitivity analysis with multiply imputed data found that the intervention and comparison groups attended a similar number of follow-up ANC visits, which could suggest that the lower attendance found in the intervention group in the main analysis could have been because of bias introduced by the missing demographic data.

Finally, the surveys were completed by a relatively small number (n=7) of midwives who were directly involved in the program, and therefore may not be representative of the views of all clinic staff involved in ANC. Future research should also collect feedback directly from pregnant women participating in the intervention to identify other areas for improvement.

Conclusions

When combined with the other limited findings available on SMS text messaging interventions for ANC, the level of interaction between women and the program may explain differences in the effectiveness of interventions. More information will not necessarily increase care-seeking behavior—it could deter women from attending antenatal visits. This intervention was relatively low intensity and likely was not sufficient to overcome larger barriers to women seeking ANC, such as transportation, inconvenience, competing priorities, and cultural factors. Further study of the specific features of SMS text messaging programs for pregnant women that contribute to their effectiveness should be a high research priority.

Acknowledgments

The authors would like to thank the Samoa Ministry of Health and National Health Service for making this study possible and the women who participated in it. The authors would also like to acknowledge the Fulbright Public Policy Fellowship, as part of this study was completed during the fellowship.

Conflicts of Interest

None declared.

References

Abbreviations

ANC: antenatal care
IRR: incidence rate ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 15.08.19; peer-reviewed by M Amoakoh-Coleman, T Esterhuizen; comments to author 24.09.19; revised version received 22.02.20; accepted 12.03.20; published 02.06.20.

Please cite as:

Watterson JL, Castaneda D, Catalani C
Promoting Antenatal Care Attendance Through a Text Messaging Intervention in Samoa: Quasi-Experimental Study
JMIR Mhealth Uhealth 2020;8(6):e15890
URL: https://mhealth.jmir.org/2020/6/e15890
doi: 10.2196/15890
PMID: 32484446
Smartphone-Guided Algorithms for Use by Community Volunteers to Screen and Refer People With Eye Problems in Trans Nzoia County, Kenya: Development and Validation Study

Hillary Rono1,2, MD; Andrew Bastawrous1,3, PhD; David Macleod4, PhD; Cosmas Bunywera2,3, BSc; Ronald Mamboleo2, BSc; Emmanuel Wanjala2, MD; Matthew Burton1,5, PhD

1Clinical Research Department, International Centre for Eye Health, London School of Hygiene & Tropical Medicine, London, United Kingdom
2Kitale County and Referral Hospital, Kitale, Kenya
3Peek Vision Foundation, London, United Kingdom
4Medical Research Council Tropical Epidemiology Group, London School of Hygiene & Tropical Medicine, London, United Kingdom
5Moorfields Eye Hospital NHS Trust, London, United Kingdom

Corresponding Author:
Hillary Rono, MD
Clinical Research Department
International Centre for Eye Health
London School of Hygiene & Tropical Medicine
Keppel Street
London
United Kingdom
Phone: 254 722627207
Email: hillary.rono@lshtm.ac.uk

Abstract

Background: The provision of eye care services is currently insufficient to meet the requirements of eye care. Many people remain unnecessarily visually impaired or at risk of becoming so because of treatable or preventable eye conditions. A lack of access and awareness of services is, in large part, a key barrier to handle this unmet need.

Objective: This study aimed to assess whether utilizing novel smartphone-based clinical algorithms can task-shift eye screening to community volunteers (CVs) to accurately identify and refer patients to primary eye care services. In particular, we developed the Peek Community Screening app and assessed its validity in making referral decisions for patients with eye problems.

Methods: We developed a smartphone-based clinical algorithm (the Peek Community Screening app) using age, distance vision, near vision, and pain as referral criteria. We then compared CVs’ referral decisions using this app with those made by an experienced ophthalmic clinical officer (OCO), which was the reference standard. The same participants were assessed by a trained CV using the app and by an OCO using standard outreach equipment. The outcome was the proportion of all decisions that were correct when compared with that of the OCO.

Results: The required sensitivity and specificity for the Peek Community Screening app were achieved after seven iterations. In the seventh iteration, the OCO identified referable eye problems in 65.9% (378/574) of the participants. CVs correctly identified 344 of 378 (sensitivity 91.0%; 95% CI 87.7%-93.7%) of the cases and correctly identified 153 of 196 (specificity 78.1%; 95% CI 71.6%-83.6%) cases as not having a referable eye problem. The positive predictive value was 88.9% (95% CI 85.3%-91.8%), and the negative predictive value was 81.8% (95% CI 75.5%-87.1%).

Conclusions: Development of such an algorithm is feasible; however, it requires considerable effort and resources. CVs can accurately use the Peek Community Screening app to identify and refer people with eye problems. An iterative design process is necessary to ensure validity in the local context.

(JMIR Mhealth Uhealth 2020;8(6):e16345) doi:10.2196/16345

KEYWORDS
visual impairment; algorithms; mobile phone; screening; mHealth; sensitivity; specificity
Introduction

Background

It is estimated that 216.6 million people globally are visually impaired (visual acuity in the better eye <6/18), and 36 million are blind (visual acuity in the better eye <3/60) [1]; about 90% of them live in low- and middle-income countries [2]. In sub-Saharan Africa, about 26 million people are visually impaired, and almost 6 million are blind [3].

The high prevalence of visual impairment (VI) is attributed to poverty and lack of access to eye services [4], shortages of health workers trained in eye care [5], and lack of awareness of the eye conditions they have [6]. Few countries in sub-Saharan Africa have reached the World Health Organization (WHO)–suggested ophthalmic cadre minimum targets of one ophthalmologist for 250,000 people to meet the surgical needs of population [7,8]. Some countries, especially in Africa, have trained midlevel personnel, including ophthalmic nurses and ophthalmic clinical officers (OCOs), to share key tasks and to compensate for the lack of ophthalmologists [9,10]. In those countries, they provide the bulk of eye care (including preventive, diagnostic, and referral services) in most rural and remote areas [11]. Generally, the few available eye health workers are concentrated in urban areas, further increasing the inequality in access to eye health care [7,12]. For example, in Trans Nzoia, a rural county in Kenya, with a population of 818,757 [13], the doctor to population ratio is 5.4 per 100,000, and the nurse to population ratio is 47 per 100,000 people [14]. This is lower than the recommended WHO minimum ratio of 230 per 100,000 population for any cadre [15].

An important strategy to improve access to eye care is task shifting, with redistribution of tasks within the health workforce, through clear referral criteria and management plans [16]. For example, guided task shifting through clinical algorithms defined as a text (flow chart) representing clinical decisions for guiding patient care [17] are a core part of the Integrated Management of Childhood Illness (IMCI) [18]. IMCI algorithms are effective in identifying pneumonia, gastroenteritis, measles, malaria, and malnutrition; however, eye conditions were not included [19]. Clinical algorithms have also been developed for use in eye care, although the accuracy of these algorithms has been variable. These include the Edinburgh Red Eye Diagnostic Algorithm to determine the correct ophthalmic diagnosis in a hospital by non–eye care nurses [20], and the Edinburgh Visual Loss Algorithm to assess the cause of visual loss by clinicians with no experience in ophthalmology [21]. Recently, the WHO developed and published clinical algorithms for primary health care (PHC) workers in Africa to assess patients with eye conditions; if proved acceptable, these algorithms could improve decision making at the PHC level [22].

Mobile health (mHealth) defined as the use of mobile and wireless technologies to support the achievement of health objectives is increasing and gaining acceptance [23,24]. There are a growing number of mHealth interventions for eye care. These include Peek Acuity, a smartphone or tablet app for measuring visual acuity [25]. A trial in primary schools in Kenya demonstrated that teachers could use Peek Acuity to detect VI (visual acuity <6/12) in school children who were aged 6 years or older [26]. This provided evidence that mHealth solutions could enable task shifting and improve access to eye health services.

In this study, we describe the process of developing and testing the Peek Community Screening app. A smartphone-based referral decision support algorithm designed to guide users to identify eye problems, which need referral using common eye signs and symptoms. To our knowledge, this is the first smartphone-based algorithm to aid referral of patients with eye problems from the community to primary eye care.

The target system users were community volunteers (CVs)—individuals who live in the community—and are selected by the community to represent them on issues of health [27]. Their roles include health promotion, referring cases to the nearest health facility, visiting homes to determine health status, and communicating with household members [28,29]. They receive a short defined informal training that is relevant to their work.

Most studies have used ophthalmologists as the reference standard [20-22]. OCOs have also been used in other studies [26,30]. In some countries where there are few ophthalmologists, OCOs provide most eye care services especially in rural areas [11]. On this basis, assessments by OCOs are acceptable. We chose OCOs because the majority of them work in rural areas (context where the app is used), they are the first contact for people with eye problems, and they have the relevant experience to make diagnoses and treatment decisions using available equipment in outreach settings. We developed a theoretical framework for assessing eye problems using principles from a framework used to train CVs to identify stroke in Pakistan (Figure 1) [31].
Objectives
The aim of this study was to develop the Peek Community Screening app and assess its validity in making referral decisions for patients with eye problems. This paper outlines the development process and the results of using the app over a number of iterations, where the algorithm was altered to improve its performance, before settling on a final algorithm to be taken forward. We describe in detail the results for the final algorithm.

Methods

Ethics Approval
The approval was granted by the London School of Hygiene and Tropical Medicine Ethics Committee, the United Kingdom, and the Institutional Research and Ethics Committee in Moi University, Eldoret, Kenya. The study adhered to provisions of the Helsinki Declaration. Written informed consent was obtained from all participants.

Development and Prevalidation Testing
We initially adopted the signs and symptoms used in a study that predicted eye conditions requiring referral in Rwanda, Madagascar, and Malawi [30], and incorporated the process used in developing the WHO clinical algorithms for PHC as a starting point for the design of our algorithms [22]. We adapted them to the environment and context for Trans Nzoia County for which the algorithms were to be used. The factors considered in making referral decisions were age, the presence of signs and symptoms of common eye problems, and visual acuity. Initially, decision trees were drawn on paper and tested informally on a small number of individuals in a hospital setting. In early tests, we observed low specificity, and incrementally changed the algorithm based on the observed results and clinical knowledge of the study authors.

From this formative work, we then developed guided questions and assessments for the CVs in order for them to be able to make referral decisions. Using the potential responses to the questions, we developed a workflow and decision matrix that were, then, translated into a digital-guided form operated on Android (Google LLC, Mountain View, CA) smartphones or tablets. The decision matrices (algorithms) were coded into a prototype app, the Peek Community Screening app, in collaboration with Peek Vision (London, UK) for use by the CVs.

We adopted a two-phase (hospital and community) prevalidation process to ensure that the final algorithm was accurate, relevant, and acceptable in this setting, and also to prepare the team adequately before the formal validation study [32]. On the basis of the clinical experience of the authors, we set the sensitivity of the algorithm to be no less than 90% and specificity above 75%. We selected and trained the CVs before commencing the prevalidation in the community setting.

Four CVs were purposefully selected from a pool of practicing CVs. A 3-day training of CVs, on how to use the Peek Community Screening app to identify and refer participants with eye problems, was conducted by two authors. Written guides, roleplays, and supervised practice sessions using consenting patients from the eye department were used for teaching purposes. Two CVs discontinued the training because of personal reasons while the remaining two CVs conducted all the validations.

To assess the consistency of CVs using the app, the same patients were independently examined by the lead author and by the two remaining CVs, all using the Peek Community Screening app to make an automated referral decision. We compared the referral decisions of the CVs with those of the lead author using the same app on the same participants.
Interrater agreement was assessed using the kappa statistic. A kappa value of 0.41 to 0.60 indicated moderate, 0.61 to 0.80 fair, and 0.81 or more indicated a good agreement [33].

We first tested the app and refined its algorithm in a hospital setting where people with a variety of eye conditions were available. We examined both the patients and their escorts (without eye problems). The purpose was to assess if the algorithm was able to identify referable eye conditions and to refine the procedures that would be followed by CVs during screening.

Following the initial hospital-based testing, we transferred the testing and refinement of the algorithms to a community setting where they would eventually be used in practice. The aim was to assess the usability of the app in identifying people with eye problems and to determine whether the target sensitivity and specificity thresholds could be met.

Interim analysis was conducted after two field tests to determine whether the target sensitivity and specificity had been achieved. For this, we compared referral decisions of the CVs using the app with that of the ophthalmologist as a reference standard. If the target sensitivity and specificity were both not met, data on the decision trees were assessed to determine which specific inputs (questions, measures, or dependencies) needed to be amended, and we made such amendments using our clinical knowledge. The changes were implemented in software, and the validation process was repeated until the sensitivity and specificity targets were met. The accepted end point was determined to be either the targets being met or when all practical combinations had been exhausted.

Validation Study

Study Design and Setting

The validation study was conducted during outreach clinics in selected communities of Trans Nzoia County, Kenya. Most outreach clinics were conducted after church services to provide a broadly representative sample from the community. All consenting participants presenting to outreach centers (irrespective of the type of illness) were eligible to participate. These participants were examined by the same CVs (who had participated in the pretesting), using the Peek Community Screening app, and by one experienced OCO, the reference standard, using standard outreach equipment. Their referral decisions (refer or not) were compared. The study was coordinated by a team from the Kitale Eye Unit.

Index Test: Referral Decisions by Community Volunteers Using the Peek Community Screening App

In the final test algorithm, users were prompted to ask the following screening questions to the parents or guardian with a child, “Does the child have any problem with their eyes today?” or directly to participant themselves, “Do you have any discomfort or pain in your eyes today?” and “Do you have a problem with your sight when seeing far or near objects?” If the participant was 6 years or older, the app prompts the user to test distance visual acuity using the Peek Acuity app and assess near visual acuity for all people aged 40 years and older at 33 cm using the RADNER reading chart (NeuMed AG) [34]. The distance visual acuity of each eye was measured separately and recorded automatically using the Peek Acuity app [35]. If the distance visual acuity was less than 6/12 in either eye or there was the presence of any self-reported eye pain or discomfort, difficulty seeing distant or near objects, or inability to see N8 on near-vision assessment for those aged 40 years or older, the participant was referred. Any eye problem in children (aged <6 years) as reported by parents or caretakers triggered a referral (Figure 2).
Reference Standard: Referral Decisions by Ophthalmic Clinical Officer Using Standard Outreach Equipment

The reference standard was the referral decision by one OCO with 14 years of experience in ophthalmology using standard equipment for outreach. He was familiar with local customs in the setting. The outreach equipment included a Snellen 6-meter vision chart to assess distance vision, RADNER reading chart for near vision, a torch, magnifying loop, i-care contact tonometer, direct ophthalmoscope, retinoscope, trial lens set, and fluorescein stains. Standard slit lamp was not used for assessment because it is not the norm to conduct a slit lamp assessment during outreach in this setting.

Study Procedures

Consecutive participants were examined for eye problems by the CVs using the app and, then, by the OCO using standard outreach equipment. The CVs followed the assessment guide...
and examined visual acuity using the embedded Peek Acuity vision test or near vision using a card when indicated. They entered the participant’s responses in the Peek Community Screening app, where a referral decision was generated automatically. Their decisions were also automatically recorded and uploaded to a dedicated cloud server once the internet connectivity was available.

After the CVs examination, the OCO masked to the decision of the CV, took a detailed history and examined the same participants. Specific information on eye pain, eye discomfort (itching and irritation), tenderness, or eye discharge was collected; vision was assessed as outlined above. A magnifying loupe and torch were used to assess the color of the conjunctiva, the appearance of the pupil, the alignment of the participants’ eyes, the presence of eye discharge, and any lid abnormalities. Direct ophthalmoscopy was used to assess the lens, vitreous, and retina. When indicated, the cornea was assessed using fluorescein and a blue light for corneal ulcers or abrasions. Intraocular pressure was measured using the i-care tonometer. A retinoscope and trial lenses were used to assess refractive errors.

A differential diagnosis for each eye was made for the purpose of management. Recording of the diagnosis followed the Kenyan Ministry of Health classification where the eye could be normal (no eye pathology) or any of the following diagnoses: cataract, corneal scars, conjunctivitis, keratitis, uveitis, retinal disease, eyelid disease, presbyopia, other refractive error, foreign body, eye growths, eye injury, and other. The OCO selected the applicable diagnosis. All patients were treated as per the OCO’s plan. The OCO recorded their decision and treatment plan on a precoded data collection form.

**Analysis**

The primary outcome was the sensitivity and specificity of the CV assessment using the Peek Community Screening app for appropriate referral decisions, compared with the OCO’s recommendation for referral. The minimum target sensitivity was 90% and specificity 75%. Positive and negative predictive values were also estimated. Logistic regression was used to identify whether there was any association between correct decisions being made by CVs and the participants’ age and sex. This was done by using the CV’s referral decisions as the outcome variable with age and sex as exposures, and the analysis was performed separately among those classed as requiring referral or not requiring referral by the reference standard.

We calculated that a sample size of 517 participants was required to estimate a sensitivity to a precision of ±5%, assuming a sensitivity of 90% and that 30.0% (155/517) of participants require referral. Thus, we aimed to recruit this number for the final iteration of the validation.

Data for CVs were downloaded from Peek’s dedicated servers in Excel format, exported to STATA, and then, cleaned and analyzed. Information from the OCO precoded questionnaire was entered into an Excel database (Microsoft, Seattle, WA, the United States), cleaned, and exported to STATA. Data were analyzed using STATA, version 15.0 (Stata Corp. LP, College Station, TX, the United States) [36]. Age was rounded up to the nearest one year, and the diagnosis was reclassified using the International Statistical Classification of Diseases and Related Health Problems [37].

**Results**

This study was conducted between November 2016 and May 2018.

**Interrater Agreement of the Community Volunteers**

During the training of the CVs, automated referral decisions were generated by the app for 59 participants, which were used to assess interrater agreement between the reference assessor (lead author) and the CVs. The reference assessor found that 75% (44/59) of the participants required referral compared with 83% (49/59), and 85% (50/59) by CV1 and CV2, respectively. There was 84.8% agreement for referral decisions between the reference assessor and CV1 and 86.4% for CV2; with a moderate kappa of 0.55 and 0.58, respectively.

**Prevalidation of the Peek Community Screening App**

One iteration in the hospital and five iterations were tested in the community before arriving at the final version (iteration seven), which was used for the validation study. The changes introduced at each iteration stage and the test performance of the versions are shown in Table 1.
Table 1. Sensitivity and specificity of the Peek Community Screening app and the changes introduced at each iteration during validation.

<table>
<thead>
<tr>
<th>Setting, iteration, and changes introduced</th>
<th>OCR decision using the Peek Community Screening app</th>
<th>Sensitivity, % (95% CI)</th>
<th>Specificity, % (95% CI)</th>
<th>PPV, % (95% CI)</th>
<th>NPV, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hospital setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 1 (enriched sample)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ask for the presence of any eye problem (no time limit); distance VA testing not mandatory for someone with eye problem</td>
<td>Refer 117</td>
<td>99.2 (95.4-100)</td>
<td>52.4 (29.8-74.3)</td>
<td>92.1 (86.0-96.2)</td>
<td>91.7 (61.5-99.8)</td>
</tr>
<tr>
<td>Do not refer 10</td>
<td>92.1 (86.0-96.2)</td>
<td>52.4 (29.8-74.3)</td>
<td>92.1 (86.0-96.2)</td>
<td>91.7 (61.5-99.8)</td>
<td></td>
</tr>
<tr>
<td>Total 127</td>
<td>99.2 (95.4-100)</td>
<td>52.4 (29.8-74.3)</td>
<td>92.1 (86.0-96.2)</td>
<td>91.7 (61.5-99.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Community setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 2 (enriched community sample)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same question above, in outreach setting with self-selected patients; ask for the presence of any eye problem (no time limit); distance VA testing not mandatory for someone with eye problem</td>
<td>Refer 250</td>
<td>98.8 (96.6-99.8)</td>
<td>66 (51.7-85.5)</td>
<td>93.3 (89.6-96.0)</td>
<td>92.1 (78.6-98.3)</td>
</tr>
<tr>
<td>Do not refer 18</td>
<td>94.8 (89.6-96.0)</td>
<td>66 (51.7-85.5)</td>
<td>93.3 (89.6-96.0)</td>
<td>92.1 (78.6-98.3)</td>
<td></td>
</tr>
<tr>
<td>Total 268</td>
<td>98.8 (96.6-99.8)</td>
<td>66 (51.7-85.5)</td>
<td>93.3 (89.6-96.0)</td>
<td>92.1 (78.6-98.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced mandatory VA testing; ask for the presence of any eye problem (no time limit); mandatory distance VA testing</td>
<td>Refer 110</td>
<td>97.3 (92.4-99.4)</td>
<td>17.8 (10.5-27.3)</td>
<td>59.8 (52.3-66.9)</td>
<td>84.2 (60.4-96.6)</td>
</tr>
<tr>
<td>Do not refer 74</td>
<td>94.8 (89.6-96.0)</td>
<td>66 (51.7-85.5)</td>
<td>93.3 (89.6-96.0)</td>
<td>92.1 (78.6-98.3)</td>
<td></td>
</tr>
<tr>
<td>Total 184</td>
<td>97.3 (92.4-99.4)</td>
<td>17.8 (10.5-27.3)</td>
<td>59.8 (52.3-66.9)</td>
<td>84.2 (60.4-96.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited the duration of eye problem to 1 day (today); ask for the presence of eye problem today; mandatory distance VA testing</td>
<td>Refer 182</td>
<td>78.4 (72.6-83.6)</td>
<td>75.6 (67.3-82.7)</td>
<td>85 (79.6-89.5)</td>
<td>66.4 (58.3-74.0)</td>
</tr>
<tr>
<td>Do not refer 32</td>
<td>64.8 (58.3-74.0)</td>
<td>75.6 (67.3-82.7)</td>
<td>85 (79.6-89.5)</td>
<td>66.4 (58.3-74.0)</td>
<td></td>
</tr>
<tr>
<td>Total 214</td>
<td>78.4 (72.6-83.6)</td>
<td>75.6 (67.3-82.7)</td>
<td>85 (79.6-89.5)</td>
<td>66.4 (58.3-74.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced eye pain instead of eye problem limited to 1 day; mandatory distance VA testing; asked—any pain in your eyes today? asked—any problem with seeing far or near objects today?</td>
<td>Refer 144</td>
<td>83.7 (77.3-88.9)</td>
<td>61.2 (52.5-69.3)</td>
<td>72.7 (66-78.8)</td>
<td>75.2 (66.2-82.9)</td>
</tr>
<tr>
<td>Do not refer 54</td>
<td>77.3 (72.8-80.9)</td>
<td>61.2 (52.5-69.3)</td>
<td>72.7 (66-78.8)</td>
<td>75.2 (66.2-82.9)</td>
<td></td>
</tr>
<tr>
<td>Total 198</td>
<td>83.7 (77.3-88.9)</td>
<td>61.2 (52.5-69.3)</td>
<td>72.7 (66-78.8)</td>
<td>75.2 (66.2-82.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduced eye discomfort; mandatory distance VA testing; asked—any eye pain or discomfort today? asked—any problem with seeing far or near objects today?</td>
<td>Refer 342</td>
<td>90.5 (87.1-93.2)</td>
<td>63.3 (57.3-69.0)</td>
<td>77.0 (72.8-80.9)</td>
<td>83.0 (77.3-87.8)</td>
</tr>
<tr>
<td>Do not refer 102</td>
<td>87.1 (71.6-91.8)</td>
<td>63.3 (57.3-69.0)</td>
<td>77.0 (72.8-80.9)</td>
<td>83.0 (77.3-87.8)</td>
<td></td>
</tr>
<tr>
<td>Total 444</td>
<td>90.5 (87.1-93.2)</td>
<td>63.3 (57.3-69.0)</td>
<td>77.0 (72.8-80.9)</td>
<td>83.0 (77.3-87.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Iteration 7: Final algorithm</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Refer 342</td>
<td>91.0 (87.7-93.7)</td>
<td>78.1 (71.6-83.6)</td>
<td>88.9 (85.3-91.8)</td>
<td>81.8 (75.5-87.1)</td>
</tr>
</tbody>
</table>
Validation Study of the Final Peek Community Screening App

We included 574 (who had completed the OCO and CV examination and outcome data) out of the potential 607 eligible participants in the analysis of the performance of the seventh iteration of the Peek Community Screening app (Figure 3). The demographic characteristics of this group are shown in Table 2.

Eye problems that needed referral were diagnosed by the OCO (reference standard) in 65.9% (378/574) of the participants. CVs using the Peek Community Screening app correctly identified 344 out of 378 (sensitivity 91.0%; 95% CI 87.7%-93.7%) participants as having referable eye conditions and 153 out of 196 (specificity 78.1%; 95% CI 71.6%-83.6%) as not. The positive predictive value was 88.9% (95% CI 85.3%-91.8%), and the negative predictive value was 81.8% (95% CI 75.5%-87.1%).

The accuracy of the algorithm varied depending on whether question alone or objectively assessed vision was used. If we used distance visual acuity and assessed near vision for those aged 40 years or older alone, without asking any of the questions about eye pain or discomfort or the question about disturbance in vision, the sensitivity dropped to 42.1% (95% CI 37.0%-47.2%), and specificity was 98.5% (95% CI 95.6%-99.7%).

If we asked about symptoms of eye pain/discomfort and disturbance in vision, with no eye examinations, the sensitivity would be 87.6% (95% CI 83.8%-90.7%) and the specificity would be 79.1% (95% CI 72.7%-84.6%). If the strategy was to refer anyone aged 40 years or older (irrespective of visual acuity of self-reported issues) and those aged under 40 who self-reported either vision problems or eye pain/discomfort, then the estimated sensitivity would be 91.5% (95% CI 88.3%-94.1%) and the specificity would be 77% (95% CI 70.5%-82.7%).

Out of the 196 participants not referred by the OCO (without eye conditions), CVs using the app incorrectly referred (false positives) 21.9% (43/196). There was no evidence to suggest that being incorrectly referred was associated with sex (odds ratio [OR] 0.70; 95% CI 0.35-1.35; P=.31) or age (OR 1.00; 95% CI 0.97-1.03; P=.86).

Further analysis of these incorrect referrals by the CVs (false positives) showed that the reasons they had been referred were as follows: 7% (3/43) of the participants could not see 6/12 (had VI), 2% (1/43) had both VI and self-reported eye pain or discomfort, 44% (19/43) had self-reported difficulty seeing distant or near objects only, 37% (16/43) had eye pain or discomfort only, and 9% (4/43) complained of both eye pain or discomfort and difficulty seeing distant or near objects. None were because of the near-vision assessment.

Similarly, out of 378 participants who were referred by the OCO (had eye problems), CVs correctly referred 91.0% (344/378). There was evidence (P=.003) of a difference in the odds of the CV using the app referring participants by age, with the odds of being referred (if referral was required according to reference standard) higher in those aged 40 or older compared with those under 40 (OR 4.38; 95% CI 1.66-11.59). This was driven by the very high referral rate in the over 40s, with the vast majority being referred both by the OCO and the CV using the app. There was no evidence (P=.28) of a difference by sex (OR 1.47; 95% CI 0.72-3.00). Most (25/34, 74%) of the participants classified as false negatives had conjunctivitis (allergic and other; Table 3).
**Figure 3.** A Standards for the Reporting of Diagnostic Accuracy Studies flow chart for study participants. CVs: community volunteers; OCO: ophthalmic clinical officer.

**Table 2.** Age, sex, and visual status of all study participants, those referred by the ophthalmic clinical officer using standard equipment and by community volunteers using the Peek Community Screening app.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total number (N=574)</th>
<th>Referred by the OCO (N=378)</th>
<th>Referred using the app (N=387)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>213 (37.1)</td>
<td>135 (65.4)</td>
<td>140 (65.7)</td>
</tr>
<tr>
<td>Female</td>
<td>361 (62.9)</td>
<td>243 (67.3)</td>
<td>247 (68.4)</td>
</tr>
<tr>
<td>Age group, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;15</td>
<td>252 (43.9)</td>
<td>128 (50.8)</td>
<td>141 (55.0)</td>
</tr>
<tr>
<td>15-29</td>
<td>100 (17.4)</td>
<td>53 (53.0)</td>
<td>55 (55.0)</td>
</tr>
<tr>
<td>30-44</td>
<td>80 (13.9)</td>
<td>57 (71)</td>
<td>53 (66)</td>
</tr>
<tr>
<td>45-59</td>
<td>76 (13.2)</td>
<td>75 (99)</td>
<td>72 (95)</td>
</tr>
<tr>
<td>60-74</td>
<td>52 (9.1)</td>
<td>51 (98)</td>
<td>52 (100)</td>
</tr>
<tr>
<td>75+</td>
<td>14 (2.4)</td>
<td>14 (100)</td>
<td>14 (100)</td>
</tr>
<tr>
<td>Visual acuity (reference), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children (vision not assessed)</td>
<td>82 (14.3)</td>
<td>41 (50)</td>
<td>40 (49)</td>
</tr>
<tr>
<td>6/6-6/12</td>
<td>411 (71.6)</td>
<td>256 (62.3)</td>
<td>268 (65.2)</td>
</tr>
<tr>
<td>6/18-6/60</td>
<td>59 (10.3)</td>
<td>59 (100)</td>
<td>57 (97)</td>
</tr>
<tr>
<td>&lt;6/60</td>
<td>22 (3.8)</td>
<td>22 (100)</td>
<td>22 (100)</td>
</tr>
</tbody>
</table>

*The distribution of the characteristics of the study participants.*

*OCO: ophthalmic clinical officer.*

*Proportions within each characteristic group that were referred by the OCO or community volunteers using the Peek Community Screening app.*
Table 3. Clinical diagnosis of the participants referred by the ophthalmic clinical officer and referral decisions by community volunteers using the Peek Community Screening app.

<table>
<thead>
<tr>
<th>Summary of diagnosis</th>
<th>Referral decision by community volunteers using the app</th>
<th>Not referred (false negatives; N=34), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Referred (N=344), n (%)</td>
<td>Not referred (false negatives; N=34), n (%)</td>
</tr>
<tr>
<td>Cataract</td>
<td>29 (8.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Presbyopia</td>
<td>56 (16.3)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>1 (2.9)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Refractive errors</td>
<td>64 (18.6)</td>
<td>2 (6)</td>
</tr>
<tr>
<td>Allergic conjunctivitis</td>
<td>117 (34.0)</td>
<td>16 (47)</td>
</tr>
<tr>
<td>Other conjunctivitis</td>
<td>44 (12.8)</td>
<td>9 (27)</td>
</tr>
<tr>
<td>Corneal disease</td>
<td>2 (0.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Retinal disease</td>
<td>5 (1.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Eye injury and foreign bodies</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Uveitis</td>
<td>1 (0.3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Pterygium conjunctival swellings</td>
<td>10 (2.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Chalazion and lid swellings</td>
<td>2 (0.3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Others</td>
<td>12 (3.5)</td>
<td>4 (12)</td>
</tr>
</tbody>
</table>

Discussion

Algorithms Development

We iteratively developed and validated smartphone-based algorithms used by CVs to identify and refer people with eye conditions for services from the community. The standard against which the algorithm was designed and validated was the referral decisions of a trained ophthalmic worker on the same participants. We predetermined in the study design the acceptable sensitivity and specificity levels to ensure adequate sensitivity to detect people with referable eye conditions in the community and also specificity that is enough not to overburden the system. This was determined as a sensitivity of not less than 90% and specificity not less than 75%.

Principal Findings

We found that 65.9% (378/574) of the participants enrolled in this study had a referable eye condition based on the examination using standard outreach equipment. This was higher than the prevalence of ocular morbidity found in other studies in Kenya and Rwanda, where the prevalence was 15.2% and 34%, respectively [38,39]. This is likely to be because of differences in the study populations and case definitions used by the studies. We conducted most validation rounds after church when most people could attend an eye check to get a representative sample of the community; however, this may not be an unbiased sample. The case definition for the earlier ocular morbidity study in Kenya excluded minor eye conditions such as pinguecula, which we included [39]. In the Rwanda national survey, only moderate to severe eye symptoms were included, but in our study, all symptoms irrespective of severity were considered [38]. We found that CVs could use the app with moderate interobserver agreement between them and the study ophthalmologist. The accuracy (sensitivity and specificity) of the algorithm was affected by prior duration of the symptoms, the commonality of symptoms and signs across different eye diseases, and the number of signs and symptoms used to generate algorithm. Sensitivity of the algorithm decreased (from 97.3% to 78.4%) with a corresponding increase in specificity (17.8% to 78.6%) when the duration of any eye symptoms was limited to one day from any duration (“Do you have any eye problem today?”). There was a simultaneous increase in specificity (from 61.2% to 63.3%) and sensitivity (from 83.7% to 90.5%) when the presence of pain was expanded to include eye discomfort. Finally, the introduction of near-vision assessment improved the specificity (from 63.3% to 78.1%). It appears that if more signs and symptoms were included in the development of the algorithm, the accuracy could be improved, but the decision to include additional elements had to be balanced with the extra cost of equipment to be used and the level of education and subsequent training requirement of CVs. Overall, the algorithm had to be accurate, acceptable, affordable, and reproducible.

Trained CVs could use the final algorithm to accurately identify and refer people with eye problems (sensitivity 91.0%) and also those without eye disease (specificity 78.1%) in the community. We observed that subjective questions were likely to cause greater variation in responses and, hence, performance of the algorithm.

For example, analysis of the referral criteria used in the algorithm show that self-reported symptoms contributed more to the sensitivity of the algorithm than objective measurement of vision. Had we not asked any of the questions on eye pain or discomfort and the one on disturbance in vision, our sensitivity would have dropped to 42.1%. This would result in missing 219 out of 378 cases determined to be those needing...
referral instead of the 34 we miss now. In fact, it would be a far better screening test to not do any eye tests at all and just ask for symptoms of eye pain or discomfort in vision. This would give us a sensitivity of 87.6% and specificity of 79.1%. Had we just asked the two questions and age, then referred anyone over 40 or who answered yes to either question, we would have got an estimated sensitivity of 91.5% and specificity of 77.0%. The findings suggest that had we excluded the objective measurement, we would have not achieved an acceptable algorithm, unless we had referred everyone older than 40 years. A population-based study in Tanzania found the prevalence of presbyopia among people aged 40 years or older to be 61.7% [40], implying that by referring everyone over 40 years, we could overload the system with false referrals. This concurs with our observation in which participants aged 40 years or older were more likely to be referred by a CV and not by the OCO (false positives).

Similarly, the same self-reported symptoms of eye pain or discomfort and self-reported poor sight contributed to inaccurate decisions from the algorithm. About 81.4% of false positive referrals using the app were from participants self-reporting to have eye discomfort or poor eyesight, whereas only 7% of false positives were because of inaccurate vision assessment. The findings suggest the need for training of the CVs to have skills in basic history taking and examinations. To reduce these false positive referrals, more clinical practice during training could improve CVs’ skills in assessing patients with eye problems. Some studies on performance of CVs [41] suggest a thorough initial training with supportive supervision to improve agreement between assessors. This implies that successful training could aim at certifying CVs who attained minimum agreement (moderate to almost perfect agreement with the reference assessor) before screening the community for eye problems. A further suggestion would be to retrain or even discontinue CVs who do not achieve the desired agreement and include a systematic way to provide continuous assessment on referral appropriateness to maintain posttraining standards.

We found that the participants who were referred by the OCO but not by the CV (false negatives) mostly (25/34, 74%) had ocular surface inflammatory conditions such as allergic conjunctivitis, presbyopia (2/34, 6%), or refractive errors (2/34, 6%; Table 3). We found that most participants with allergic conjunctivitis were correctly referred, suggesting that those identified as false negatives, may have had mild symptoms. This could have resulted from self-reported symptoms that were selectively mentioned to the CV but not to the OCO. Although we did not analyze the severity of allergic conjunctivitis to conclusively classify them as false negatives, other studies have found that some patients who presented with red eyes and allergic conjunctivitis for outpatient consultations had less severe conjunctivitis that could be transient or managed at primary point of contact [42,43].

Future Improvements

The findings, therefore, suggest the need for a deeper understanding and analysis of allergic eye conditions according to severity. There are suggestions to improve the sensitivity of current algorithm. The first approach is to introduce an assessment for red eyes into the algorithm with integrated images of different types of red eyes to aid in the classification of severity. The second approach is upscaling screeners’ knowledge to distinguish normal and allergic eye disease. The ideal CVs should, therefore, have the skill set to identify VI, referable and nonreferable allergy, and Identification and management of presbyopia. This could, however, require policy change to implement in practice.

Finally, it may be possible to recalibrate the referral criteria for VI based on the capacity of the services, restricting the threshold of referrals to a level that generates referrals of those with more severe VI and lowering this threshold over time as capacity increases to ensure the health system is not overburdened.

As demonstrated, there are multiple factors that affect the performance and acceptance of a guided screening algorithm. These include the subjective and objective inputs in the decision tree. Objective threshold tests such as acuity lead to a binary output (pass or fail), whereas subjective assessments such as self-perception of vision loss have a spectrum of outputs that requires a binary threshold to be derived to progress through the decision tree. Every iteration requires a significant amount of time and resource, making optimization challenging in practice. There is a potential for utilizing Web-based A/B testing techniques currently being used in digital marketing to optimize algorithms more rapidly [44].

Limitations

There are limitations to be considered in this study. The study was conducted after church services and could have excluded those who did not attend church. Moreover, those who participated may have had a perceived eye problem, which could have resulted in higher prevalence of referable eye conditions and, hence, higher predictive values. There could also be diagnostic uncertainty in the reference standard in this study where an OCO used simple outreach equipment without a slit lamp. The OCOs used as the reference are not available in other health systems and, therefore, the results may be not generalizable to those setting.

Conclusions

The Peek Community Screening app meets the minimum predetermined criteria. The next step is to incorporate the algorithm into a screening system to assess performance in a health system, to identify people with eye problems, and to link them to primary and secondary centers. We anticipate that more people with eye health needs will be able to access the appropriate level of eye services. More validation studies conducted in different settings and improvement to the existing algorithm may be required. Further research on the performance of the algorithm is needed for specific ages groups (aged 15 years or less, 15-40 years, and those 40 year and older). If acceptable standards are met, it could be of value in determining the population demand for eye services in population-based studies and for being a validated methodology for increasing access to appropriate services in integrated eye health programs.
Acknowledgments

The research was funded through the Commonwealth Eye Health Consortium (CEHC), with funding from the Queen Elizabeth Diamond Jubilee Trust’s Avoidable Blindness program. The CEHC is administered through the International Centre for Eye Health based at the London School of Hygiene & Tropical Medicine. MB is supported by the Wellcome Trust 207472/Z/17/Z. The funders had no role in the design, execution, interpretation, or writing of the study. The authors would like to thank Kitale county and referral hospital eye unit staff, Kenya, for their support. This submission has not been published anywhere previously and is not simultaneously being considered for any other publication.

Authors’ Contributions

HR and AB conducted the literature search. HR, AB, DM, and MB contributed to the study conception, design, and methodology. CB, HR, and RM collected the data. HR and DM performed the statistical analysis. HR drafted the manuscript. All authors are responsible for the critical revision of the manuscript for important intellectual content. AB and MB obtained the funding. EW and CB provided the administrative, technical, or material support. AB and MB supervised the study.

Conflicts of Interest

The Peek Vision Foundation (09919543) is a registered charity in England and Wales (1165960) with a wholly owned trading subsidiary, Peek Vision Ltd (09937174). MB is a Trustee of The Peek Vision Foundation, and AB is the chief executive officer of The Peek Vision Foundation and Peek Vision Ltd. CB works at Peek Vision Ltd. HR is an advisor to Peek Vision Ltd. All other authors declare no conflicts of interest.

References


36. Stata: Software for Statistics and Data Science. College Station, TX: StataCorp LP URL: https://www.stata.com/ [accessed 2020-02-05]


Abbreviations

CEHC: Commonwealth Eye Health Consortium
CV: community volunteer
IMCI: Integrated Management of Childhood Illness
mHealth: mobile health
OCO: ophthalmic clinical officer
OR: odds ratio
PHC: primary health care
VI: visual impairment
WHO: World Health Organization

©Hillary Rono, Andrew Bastawrous, David Macleod, Cosmas Bunywera, Ronald Mamboleo, Emmanuel Wanjala, Matthew Burton. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 19.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
The Clinical and Cost-Effectiveness of Telerehabilitation for People With Nonspecific Chronic Low Back Pain: Randomized Controlled Trial

Francis Fatoye¹, BSc, MSc, MBA, PhD; Tadesse Gebrye¹, BSc, MSc, MPH; Clara Fatoye¹, BSc, MA, PGD; Chidozie E Mbada², BSc, MSc, PhD; Mistura I Olaoye², BSc, MSc; Adesola C Odole³, BSc, MSc, PhD; Olumide Dada³, BSc, PhD

¹Faculty of Health, Psychology and Social Care, Manchester Metropolitan University, Manchester, United Kingdom
²College of Health Sciences, Obafemi Awolowo University, Osun, Nigeria
³College of Medicine, University of Ibadan, Oyo, Nigeria

Corresponding Author:
Francis Fatoye, BSc, MSc, MBA, PhD
Faculty of Health, Psychology and Social Care
Manchester Metropolitan University
Birley Fields Campus, Bonsall Street
Manchester, M15 6GX
United Kingdom
Phone: 44 0161 247 2963
Email: f.fatoye@mmu.ac.uk

Abstract

Background: Telerehabilitation can facilitate multidisciplinary management for people with nonspecific chronic low back pain (NCLBP). It provides health care access to individuals who are physically and economically disadvantaged.

Objective: This study aimed to evaluate the clinical and cost-effectiveness of telerehabilitation compared with a clinic-based intervention for people with NCLBP in Nigeria.

Methods: A cost-utility analysis alongside a randomized controlled trial from a health care perspective was conducted. Patients with NCLBP were assigned to either telerehabilitation-based McKenzie therapy (TBMT) or clinic-based McKenzie therapy (CBMT). Interventions were carried out 3 times weekly for a period of 8 weeks. Patients' level of disability was measured using the Oswestry Disability Index (ODI) at baseline, week 4, and week 8. To estimate the health-related quality of life of the patients, the ODI was mapped to the short-form six dimensions instrument to generate quality-adjusted life years (QALYs). Health care resource use and costs were assessed based on the McKenzie extension protocol in Nigeria in 2019. Descriptive and inferential data analyses were also performed to assess the clinical effectiveness of the interventions. Bootstrapping was conducted to generate the point estimate of the incremental cost-effectiveness ratio (ICER).

Results: A total of 47 patients (TBMT, n=21 and CBMT, n=26), with a mean age of 47 (SD 11.6) years for telerehabilitation and 50 (SD 10.7) years for the clinic-based intervention, participated in this study. The mean cost estimates of TBMT and CBMT interventions per person were 22,200 naira (US $61.7) and 38,200 naira (US $106), respectively. QALY gained was 0.085 for TBMT and 0.084 for CBMT. The TBMT arm was associated with an additional 0.001 QALY (95% CI 0.001 to 0.002) per participant compared with the CBMT arm. Thus, the ICER showed that the TBMT arm was less costly and more effective than the CBMT arm.

Conclusions: The findings of the study suggested that telerehabilitation for people with NCLBP was cost saving. Given the small number of participants in this study, further examination of effects and costs of the interventions is needed within a larger sample size. In addition, future studies are required to assess the cost-effectiveness of this intervention in the long term from the patient and societal perspective.

(JMIR Mhealth Uhealth 2020;8(6):e15375) doi:10.2196/15375

KEYWORDS
cost-utility analysis; quality-adjusted life years; telerehabilitation; low back pain; mobile phone
Introduction

Low back pain (LBP) can result from several different abnormalities or diseases. It is commonly accompanied by pain in one or both legs, between the lower rib margins, and in the buttock creases [1]. Almost 90% and 10% cases of LBP are of nonspecific and specific causes, respectively [2]. The prevalence of LBP in those aged 9 to 18 years in high-income, medium-income, and low-income countries was around 40.0% [3]. It has also been reported that most adults will have LBP at some point during their lifetime [4]. LBP was responsible for around 60.1 million years lived with disability globally in 2015, and there will be an overall increase in its global burden because of population increase and aging [5]. The working age groups in middle-income and low-income countries have the highest disability from LBP [6]. A review of studies in the United States and internationally suggested that the costs of treating LBP are extremely high, where indirect costs represented a majority of the overall costs associated with LBP [7]. Dagenais et al [7] also indicated that the largest proportion of direct medical costs for LBP was spent on physical therapy and inpatient hospital services, followed by pharmacy and primary care. In relation to nonspecific chronic low back pain (NCLBP), there are no specific treatments that can be provided. The reason for this is that the pathoanatomical cause for nonspecific LBP is unknown [8].

Many clinical practice guidelines are recommended for the prevention and management of LBP [9]. These practice guidelines include education that supports self-management and resumption of normal activities and exercise, use of medication, imaging, and surgery. Research studies from high-income countries suggest that exercise alone, and exercise in combination with education, reduces the risks of an episode of LBP [10]. Compared with no treatment, a supervised exercise for children and adolescents can improve average pain intensity by 2.9 points (95% CI 1.6 to 4.1) in patients with LBP [11]. On the other hand, Steffens et al [10] concluded that physiotherapy interventions such as education alone, back belts, and shoe insoles did not appear to prevent LBP.

Despite the availability of many clinical guidelines for managing LBP, a substantial difference in their applicability exists in high-income as well as low-income and middle-income countries [12]. Identifying the best intervention for LBP can not only improve the health outcomes for patients but also reduce health care utilization and costs associated with the management of the condition. Telerehabilitation, in the form of a mobile phone app platform extension exercise that enables a patient to perform exercises using a smartphone, may be a practical intervention for LBP in geographically remote areas with a shortage of services and a lack of access to physical therapy rehabilitation services. Telerehabilitation uses communication technology for the remote delivery of care to patients and has the potential to manage multiple components of health, including functional independence, self-care, and self-management of illness [13].

The findings from a review of 29 articles indicated that telehealth had a moderate, positive, and significant effect on clinical outcomes for different patient populations, including LBP, heart, and psychiatric conditions [14]. In a few studies included in the systematic review, the use of telerehabilitation for patients with LBP was reported to have positive clinical outcomes, which may in turn lead to fewer visits to the emergency room and physician, fewer admission to hospitals, shorter length of stay in hospitals, and lower costs [14]. Despite the methodological differences in studies and the health care system of various countries, understanding the clinical outcomes and the economic costs of telerehabilitation interventions may improve their efficiency. The use of telerehabilitation in low- and middle-income countries such as Nigeria is just emerging; as a result, data on the clinical and cost-effectiveness of telerehabilitation are scarce [15,16]. To date, we are not aware of any study that has investigated the clinical and cost-effectiveness of physiotherapy using telerehabilitation in these countries. To study the clinical and cost-effectiveness of telerehabilitation, we developed a telerehabilitation-based McKenzie exercise intervention for people with NCLBP. This study, therefore, assessed the clinical and cost-effectiveness of telerehabilitation-based McKenzie therapy (TBMT) compared with clinic-based McKenzie therapy (CBMT) for people with NCLBP in Nigeria.

Methods

Trial Design

This study was an experimental research design and was conducted at the department of physiotherapy, Ladoke Akintola University of Technology University Teaching Hospital, Osogbo, and the physiotherapy department, State Hospital, Ejigbo. Ethical approval for this study was obtained from the Health Research Ethical Committee of the Institute of Public Health, Obafemi Awolowo University (registration number: IPH/OAU/12/515).

Study Population

The sample size for this study was determined using equation 1 [17]:

$$m = \frac{\pi_1 (1-\pi_1) + \pi_2 (1-\pi_2)}{\pi_1 - \pi_2}$$

(1)

where $c=7.9$ for 80% power and $\pi_1$ and $\pi_2$ are the proportion estimates ($\pi_1=0.25$ and $\pi_2=0.65$). Therefore, $m=0.25 [(1-0.25) + 0.65 (1-0.65)]/(0.25-0.65)^2=20.49$, which is approximately 21. Hence, the calculated sample size was 42 (21 per group). To account for a possible attrition of 10% (ie, 4.2), the estimated minimum sample size was 46.

Patients with NCLBP, who attended outpatient physiotherapy departments, were recruited into this study. At the start of the recruitment process, the purpose of the research was explained to the participants. All participants (n=70) who were assessed for eligibility in the study were provided an informed written consent form translated by experts into the local language. A research assistant recorded the number of participants who were invited to participate, the number of participants who declined to participate, and the number of screened patients who were not eligible and their reasons for declining participation.
Eligibility for participation in this study was based on physician referral and physiotherapists’ diagnosis of NCLBP. Participants with a clinical diagnosis of long-term NCLBP aged between 20 and 65 years and those without any obvious deformities affecting the trunk or upper and lower extremities were included. The term long-term was used in this study instead of chronic. Using the International Classification of Functioning, Health and Disability framework, it is believed that the word chronic may be associated with negative expectations; therefore, the word long-term is preferred [18]. In addition, patients included in the study were those without any apparent deformities in the trunk and upper and lower extremities. To have a homogeneous sample of LBP type that is amenable to the McKenzie therapy, directional preference for extension was a major inclusion criterion. Directional preference is defined as the movement or posture that decreases or centralizes pain that emanates from the spine or increases the range of movement [19]. Patients with LBP who had a known comorbidity or history of cardiovascular disease for which exercise was contraindicated were excluded from this study. In addition, patients who were pregnant, those who had a previous back surgery or an experience of the McKenzie therapy, and those with directional preference for flexion or no directional preference based on the McKenzie assessment were excluded from this study.

Randomization
A research assistant who was not involved in the assessment and treatment of the participants randomly allocated participants to the different treatment groups. The same assistant who was not involved in the assessment and treatment of the participants randomly allocated participants who volunteered to participate and satisfied the eligibility criteria to the different treatment groups (A or B). To ensure equal-sized treatment groups, random permuted blocks were used [20], and a block size of 4 was chosen (ie, AABB, ABAB, and all the other possible restricted permutations). The block permutations were computer generated using a factorial equation formula shown in equation 2:

\[
\frac{(4!)}{(2!)(2!)} = 24 \text{ (2)}
\]

The consecutive participants were randomized following the computer-generated block permutations. The printouts of the 24 restricted computer-generated block permutation sequences were sequentially numbered, cut, and placed in a sealed envelope.

This study utilized blocked randomization because of its advantage to ensure an equal-size treatment group. Hence, this rigorous assignment method was intended to be a strength to the design of the study. However, the differences in sample size between groups were not due to random assignment but because participants declined or refused to participate, which was beyond the control of the researchers. The participants were randomly assigned to either the CBMT group or the TBMT group.

Telerehabilitation-Based McKenzie Therapy
The TBMT group received a mobile phone–based app of Mechanical Diagnosis and Therapy (MDT). Most of the participants in the TBMT group were provided with smartphones within the available budget. Others with their own phones were recruited into that arm of the study to be able to achieve a minimum sample size, whereas those without an Android phone that could run the app were excluded.

TBMT is a comparable version of CBMT performed at home with the assistance of a mobile phone app. The mobile app is a combination of the McKenzie extension protocol and back care education developed and enabled to run on a smartphone or an Android phone with an operating system of version 3.5. TBMT is a mobile phone video app designed for patients with chronic LBP. The app incorporated personalized and guided self-therapy using the same protocol as the McKenzie protocol (ie, extension lying prone, extension in prone, and extension in standing). Performance feedback and progress tracking were telemonitored through enhanced caregiver support to improve patient engagement and therapy compliance.

Clinic-Based McKenzie Therapy
The CBMT group received the McKenzie extension protocol and a set of back care education instructions comprising a 9-item instructional guide on standing, sitting, lifting, and other activities of daily living at home [19]. The protocol involves a course on specific lumbosacral repeated movements in extension that cause the symptoms to centralize, decrease, or abolish [21]. The extension activities include extension lying prone, extension in prone, and extension in standing repeated up to 10 times [19,21]. The determination of the directional performance for extension was followed by the extension protocol. The details of the protocol have been described in an earlier publication [22].

- Extension lying prone: participant laid prone, with elbows placed under the shoulders so that he/she could lean on the forearms, and stayed in this position for 5 min. The movement was repeated up to 10 times.
- Extension in prone: participant positioned in prone, placed his/her hands under the shoulders in the press-up position. The participant then straightened the elbows and pushed the top half of the body up as far as his/her pain permits. The participant maintained the position for up to 2 seconds. The movement was repeated up to 10 times.
- Extension in standing: participant stood upright with the feet slightly apart and placed his/her hands in the small of the back with the fingers pointing backward. The participant then stretched the trunk backward at the waist level as far as he/she can, using the hands as a fulcrum while keeping the knees straight. The movement was repeated up to 10 times.

Outcomes and Assessment
Baseline assessment was carried out for each participant who was recruited into the study. Anthropometric variables such as weight and height were measured. Information such as age, gender, educational level, occupation, marital status, onset of back pain, recurrence, duration of complaint, and previous intervention were recorded for each participant accordingly. The participants were also assessed for directional preference. It involved repeated movements, of 5 to 10 sets of each movement, and it included movements in standing and lying positions and in sagittal and frontal planes while the participants’
symptomatic and mechanical responses were assessed. Following the repeated-movement testing, the participants returned to the same standing position, and following standardized instructions in the McKenzie Institute’s Lumbar Spine Assessment Algorithm (MILSAA), they were asked if the pain was centralizing or peripheralizing during and after movements or if there was no effect. The MILSAA is a well-defined algorithm that leads to the simple classification of spine-related disorders. This is based on a consistent cause and effect relationship between historical pain behavior as well as the pain response to repeated test movements, positions, and activities during the assessment process. The participants’ mechanical response to repeated movements was used to establish their directional preference.

Treatment health outcomes were assessed at 4 weeks and 8 weeks of the study, and the outcome evaluators were blinded to the groups and the interventions. A primary outcome of the LBP disability was used as a health outcome, which was measured by the Oswestry Disability Index (ODI). The ODI is a self-administered questionnaire on a 10-item scale with 6 response categories [18]. Each item scores from 0 (better) to 5 (worse). Each score was transferred into a 0 to 100 scale. The ODI score of each patient was recorded. To estimate the health-related quality of life of patients, the ODI score was mapped to short-form six dimensions (SF-6D) instrument using equation 3 [23]:

\[ SF-6D = 0.78275 – 0.00518 \times (ODI) \]  

The SF-6D is a preference-based health state classification system [24]. The SF-6D values obtained using the above formula were important for measuring the health outcomes of patients, and this enabled the researchers to perform a cost-utility analysis (CUA). The CUA is used to determine the cost in terms of utilities, and it combines the quantity and quality of life. An increased quality of life of LBP participants can be expressed as a utility value on a scale of 0 (dead) to 1 (perfect quality of life). After obtaining the SF-6D values of each participant, the quality-adjusted life year (QALY) of each participant was calculated. QALY was calculated by multiplying the SF-6D values and the duration of time (years). For the purpose of this study, the average of QALYs at 4 weeks and 8 weeks was considered for the participants in the study.

**Resource Use and Costs**

Health care resource use and costs were assessed based on the McKenzie extension protocol, focusing on the direct implementation of costs of TBMT and CBMT. The direct health care resources included for implementation were the back-treatment DVD that was used for dummy app development before the real app was developed, development of the mobile phone–based app of the MDT for smartphones, and Android phones with an operating system of version 3.5. In addition to these, smartphones with the app installed for patients who may not have smartphones, phone credits for calls, internet data use for the entire project period, and fee for consultations were among the resources used. These resources were documented from McKenzie therapy protocols. Personal costs associated with CBMT were not included in this analysis. As the patients were those attending outpatient physiotherapy departments, the costs of medications were not included in this study. Moreover, in the context of this study, most of the patients can access health care through out-of-pocket means, in addition to undisclosed self-medication practices that are often encouraged by over-the-counter access to more than the regulated medications.

**Statistical and Cost-Effectiveness Analysis**

Descriptive statistics of the mean or SD and an inferential data analysis were performed using Statistical Packages for the Social Sciences version 23 (IBM Corp). A nonparametric Mann-Whitney U test and Friedman test were used to compare the mean effects between the treatment regimen across the fourth- and eighth-week period and the changes of the effects of the interventions from baseline at the fourth week and eighth week for the categorical variables, respectively. A significance level of \( P = 0.05 \) was adopted for those comparisons.

The incremental cost-effectiveness ratio (ICER) was used to assess the cost-effectiveness of TBMT compared with CBMT using the formula shown in equation 4 [25]:

\[ \text{ICER} = \frac{\Delta \text{Cost}}{\Delta \text{Effectiveness}} = \frac{(\text{Cost of TBMT} - \text{Cost of CBMT})}{(\text{QALY for TBMT} - \text{QALY for CBMT})} \]  

The ICER is the differential costs and outcomes between the new intervention (TBMT) and the control (CBMT). The numerator in the cost-effectiveness ratio is the monetary cost of the TBMT intervention minus the monetary cost of CBMT. The annual costs of the projects were calculated by converting the 8-week costs, the period used for implementation. The denominator is the QALY gained by TBMT minus the QALY gained by CBMT. Bootstrapping was used for a pair-wise comparison of the mean costs and effects between the TBMT and CBMT groups. CIs for the mean differences in effects were obtained by bootstrapping (1000 replications). The bootstrapped cost and effect pairs were also graphically represented on a cost-effectiveness plane [26].

**Results**

**Data Source and Selection**

A total of 47 participants (CBMT, \( n = 26 \) and TBMT, \( n = 21 \)) were randomized and provided baseline data (Figure 1). Table 1 shows the baseline characteristics of these participants. The occupations of the participants were trading (\( n = 13 \)), teaching (\( n = 7 \)), nursing (\( n = 3 \)), tailoring (\( n = 6 \)), and others (\( n = 18 \)). The mean age of the participants was 47.3 (SD 11.6) years and 50 (SD 10.7) years for the TBMT group and CBMT group, respectively. The participants in the TBMT group had higher weight and BMI (8.1 kg and 1.5 kg/m\(^2\), respectively) than those in the CBMT group. A pain duration of 9.8 (SD 2.7) months was reported for the participants in the TBMT group, which was less than that of the CBMT group, a pain duration of 8.3 (SD 3.2) months. From this study, weight (kg) was the only anthropometric characteristic that was significantly different between groups at baseline. However, BMI was not statistically different between both groups. The most common causes of chronic LBP in the participants were lifting, poor posture, prolonged sitting, bending, standing, and rigorous activity.
Figure 1. Flowchart of included patients. CBMT: clinic-based McKenzie therapy; TBMT: telerehabilitation-based McKenzie therapy.

Table 1. Baseline characteristics of the telerehabilitation-based McKenzie therapy group and clinic-based McKenzie therapy group.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Telerehabilitation-based McKenzie therapy group (n=21)</th>
<th>Clinic-based McKenzie therapy group (n=26)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>47.3 (11.6)</td>
<td>50.0 (10.7)</td>
<td>.40</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>79.1 (13.1)</td>
<td>71.0 (7.8)</td>
<td>.01</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>27.9 (3.6)</td>
<td>26.4 (3.4)</td>
<td>.15</td>
</tr>
<tr>
<td>Height (m), mean (SD)</td>
<td>1.7 (0.1)</td>
<td>1.6 (0.1)</td>
<td>.11</td>
</tr>
<tr>
<td>Pain duration (months), mean (SD)</td>
<td>9.8 (2.7)</td>
<td>8.3 (3.2)</td>
<td>.10</td>
</tr>
</tbody>
</table>

Occupation, n

- Trading  4  9
- Teaching  2  5  N/A
- Nursing   2  1  N/A
- Tailoring 2  4  N/A
- Artisan   4  2  N/A
- Driver    0  1  N/A
- Civil service  6  4  N/A
- Student   1  0  N/A

N/A: not applicable.

Resource Use and Costs

Participants in the CBMT and TBMT groups provided the cost data (Table 2). The cost estimates for SMS messages and reminder calls were 50 naira (US $0.14) per unit, and the cost estimate of owning a compatible phone for the app was 20,000 naira (US $55.56). The resource use and costs for CBMT were as follows: cost estimate of each clinic visit (3 visits per week; 1000 naira [US $2.78] per visit) and transportation and refreshment estimate for each clinic visit (500 naira [US $1.39] per visit). Moreover, the common costs of both groups were costs of physiotherapy consultation (before randomization into the group), and they were estimated to be 1000 naira (US $2.78).
Table 2. Cost associated with implementation of telerehabilitation-based McKenzie therapy and clinic-based McKenzie therapy.

<table>
<thead>
<tr>
<th>Resources</th>
<th>Cost per visit (US$)</th>
<th>Total cost per participant (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TBMT</td>
<td>CBMT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS messages and reminder calls (3 times per week)</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Compatible phones for the app</td>
<td>55.6</td>
<td>N/A c</td>
</tr>
<tr>
<td>Clinic visit (3 visits per week)</td>
<td>N/A</td>
<td>2.8</td>
</tr>
<tr>
<td>Consultation fee</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>Transportation and refreshment</td>
<td>N/A</td>
<td>1.4</td>
</tr>
<tr>
<td>Total cost</td>
<td>N/A</td>
<td>61.8</td>
</tr>
</tbody>
</table>

aTBMT: telerehabilitation-based McKenzie therapy.
bCBMT: clinic-based McKenzie therapy.
cN/A: not applicable.

Effectiveness

The mean change of clinical effectiveness of CBMT and TBMT from baseline at weeks 4 and 8 is presented (Table 3). The changes of health outcomes from baseline to week 4 and week 8 have shown a significant difference (P<.001) within the CBMT and TBMT groups. However, no significant or clinically relevant mean ODI score difference was observed in the measurements at weeks 4 and 8 between the CBMT and TBMT groups (P>.05).

Table 3. Estimates of clinical effectiveness at weeks 4 and 8 after randomization.

<table>
<thead>
<tr>
<th>Oswestry Disability Index</th>
<th>Mean change from baseline (95% CI)a</th>
<th>Mean treatment difference (95% CI)b</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinic-based McKenzie therapy</td>
<td>Telerehabilitation-based McKenzie therapy</td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>8.5 (5.45 to 11.55)</td>
<td>10.43 (7.74 to 11.54)</td>
<td>1.61 (−2.1 to 5.43)</td>
</tr>
<tr>
<td>Week 8</td>
<td>14.50 (10.63 to 18.36)</td>
<td>15.71 (12.85 to 18.57)</td>
<td>0.81 (−2.39 to 4.01)</td>
</tr>
</tbody>
</table>

aP<.001.
bThe mean treatment difference suggests the comparison of the health outcomes of CBMT and TBMT at week 4 and week 8.

Cost-Effectiveness

Table 4 reports the point estimates of the incremental costs and effects per patient. A reduction in the total health care cost in the participants who received TBMT was reported, 16,000 naira (US $44.26), compared with those who received CBMT. On the other hand, participants who received TBMT had an additional health benefit (0.001 QALY) compared with those who received CBMT. Thus, the ICER showed that the TBMT arm was less costly and more effective than the CBMT arm. Figure 2 shows the incremental cost-effectiveness plane for a plot of 1000 bootstrap incremental cost and effect resample means.

Table 4. Incremental cost-effectiveness ratio.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Cost, naira (US$)</th>
<th>Incremental cost, naira (US$)</th>
<th>Effects, mean QALYa (95% CI),</th>
<th>Incremental effect, mean QALYb (95% CI),</th>
<th>Incremental cost-effectiveness ratio of naira (US$)/QALY gained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinic-based McKenzie therapy</td>
<td>38,200 (106.22)</td>
<td>N/A b</td>
<td>0.084 (0.084 to 0.085)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Telerehabilitation-based McKenzie therapy</td>
<td>22,200 (61.7)</td>
<td>−16,000 (−44.26)</td>
<td>0.085 (0.80 to 0.09)</td>
<td>0.001 (0.001 to 0.002)</td>
<td>Dominant</td>
</tr>
</tbody>
</table>

aQALY: quality-adjusted life year.
bN/A: not applicable.
**Discussion**

**Overview**

This is the first study to examine the clinical and cost-effectiveness of telerehabilitation compared with clinic-based therapy. The mean treatment effect of the participants was assessed at week 4 and week 8. A significant difference was found for clinical effectiveness within the TBMT and CBMT groups from baseline to week 4 and week 8. On the other hand, no significant difference of the mean ODI score was reported between the two intervention groups. The findings of this study are in line with the results of the study by Kosterink et al [27], who investigated the effects of a 4-week teletreatment service in subjects with nonspecific neck and shoulder pain, where they showed that the treatment was effective in reducing pain intensity and disability over time. They also reported that there was no significant difference between teletreatment and conventional care—where subjects did not receive any specific intervention such as osteopathy, chiropractice, ergonomic counseling, medication, physiotherapy, acupuncture, stress management, and relaxation training.

In line with the study conducted in Norway on patients with musculoskeletal problems, the results of this study indicated that telerehabilitation therapy was cost saving [28]. Also, it is understood that both cost and health benefits of the two interventions could have an impact on the cost-effectiveness of telerehabilitation. This study showed that telerehabilitation was less costly than clinic-based treatment. In line with our study, a cost analysis study in Canada, which was conducted on patients with a knee problem, also concluded that the cost of telerehabilitation was lower than that of conventional rehabilitation [29]. Moreover, a cost-minimization study on patients with skeletal problems in Finland has also indicated that telemedicine was less costly for society than conventional care at a workload of more than 80 patients per year [30].

The increment or reduction of the costs and effectiveness of the TBMT by half from the base case values was unlikely to affect its cost-effectiveness in this study. The findings of this study are consistent with the results of the cost-effectiveness analysis study on telemedicine for primary care delivery, where telemedicine was shown to be cost saving as long as its effectiveness was greater than that of the controlled intervention [28]. However, the reduction of the health benefits from the base case values in this study could lead telerehabilitation not to be a cost-effective intervention. Overall, it is important that patients adhere to telerehabilitation services and improve their health for the new intervention to be cost-effective.

TBMT was approximately 50% cheaper than CBMT; this is because of less requirements of a clinic-based facility and less contact with a physiotherapist for its delivery. In other words, there is an opportunity to implement telerehabilitation programs across numerous geographic locations if needed. In low-income countries, such as Nigeria, access to physiotherapy services is a challenge because of the shortage of physiotherapists and limited access to clinic-based programs [31]. Unlike CBMT, TBMT could overcome barriers to accessing physiotherapy services and could provide numerous benefits with reduced cost to the patients in Nigeria. However, the key challenges for its implementation strategies are the existence of effective internet services and patient reluctance to engage [32].

**Strengths and Limitations**

The major strength of this study was that it is the first study in low- and middle-income countries to evaluate the cost-effectiveness of telerehabilitation therapy for patients with NCLBP using a randomized controlled trial. In addition, the findings of this study could inform clinicians and decision makers about the implementation of TBMT as a complementary option of CBMT services in Nigeria. On the other hand, the findings reported here should be viewed in the context of the limitations of this study. The cost analysis did not include costs of medications and indirect costs. It is believed that the exclusion of costs of medications and indirect costs to the cost-effectiveness analysis may underestimate the total cost of therapies. The second limitation of the study was related to the time of follow-up; the effects of the telerehabilitation therapies might be different in the long-term follow-up. Thus, evidence of health benefits from a long-term follow-up of patients is important to be incorporated in the cost-effectiveness analysis of telerehabilitation.

**Conclusions**

The findings of this study showed that telerehabilitation was associated with greater health benefits and lower costs, suggesting that it was a cost-saving therapy compared with a clinic-based therapy. This suggests that the implementation of TBMT could help to overcome barriers to access to physiotherapy services, particularly in low-income countries such as Nigeria, thereby improving the health outcomes of patients in these countries. Future studies are required to assess...
the cost-effectiveness of the intervention in the long term from the patient and societal perspective.

Acknowledgments
This research was partially funded by an African Doctoral Dissertation Research Fellowship re-entry grant awarded by the African Population and Health Research Center in partnership with the International Development Research Centre.

Conflicts of Interest
None declared.

Editorial notice: This randomized study was not registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low. Readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).

References


Abbreviations

- CBM: clinic-based McKenzie therapy
- CUA: cost-utility analysis
- ICER: incremental cost-effectiveness ratio
- LBP: low back pain
- MDT: Mechanical Diagnosis and Therapy
- MILSA: McKenzie Institute’s Lumbar Spine Assessment Algorithm
- NCLBP: nonspecific chronic low back pain
- ODI: Oswestry Disability Index
**QALY**: quality-adjusted life year

**SF-6D**: short-form six dimensions

**TBMT**: telerehabilitation-based McKenzie therapy
Original Paper

mHealth App Prescription in Australian General Practice: Pre-Post Study

Oyungerel Byambasuren¹, MD, MMedRes; Elaine Beller¹, MAppStat; Tammy Hoffmann¹, PhD; Paul Glasziou¹, MD, PhD
Institute for Evidence-Based Healthcare, Bond University, Robina, Australia

Corresponding Author:
Oyungerel Byambasuren, MD, MMedRes
Institute for Evidence-Based Healthcare
Bond University
14 University Dr,
Robina, 4226
Australia
Phone: 61 755955518
Email: obyambas@bond.edu.au

Abstract

Background: Evidence of effectiveness of mobile health (mHealth) apps as well as their usability as non-drug interventions in primary care are emerging around the globe.

Objective: This study aimed to explore the feasibility of mHealth app prescription by general practitioners (GPs) and to evaluate the effectiveness of an implementation intervention to increase app prescription.

Methods: A single-group, before-and-after study was conducted in Australian general practice. GPs were given prescription pads for 6 mHealth apps and reported the number of prescriptions dispensed for 4 months. After the reporting of month 2, a 2-minute video of one of the apps was randomly selected and sent to each GP. Data were collected through a prestudy questionnaire, monthly electronic reporting, and end-of-study interviews. The primary outcome was the number of app prescriptions (total, monthly, per GP, and per GP per fortnight). Secondary outcomes included confidence in prescribing apps (0-5 scale), the impact of the intervention video on subsequent prescription numbers, and acceptability of the interventions.

Results: Of 40 GPs recruited, 39 commenced, and 36 completed the study. In total, 1324 app prescriptions were dispensed over 4 months. The median number of apps prescribed per GP was 30 (range 6-111 apps). The median number of apps prescribed per GP per fortnight increased from the pre-study level of 1.7 to 4.1. Confidence about prescribing apps doubled from a mean of 2 (not so confident) to 4 (very confident). App videos did not affect subsequent prescription rates substantially. Post-study interviews revealed that the intervention was highly acceptable.

Conclusions: mHealth app prescription in general practice is feasible, and our implementation intervention was effective in increasing app prescription. GPs need more tailored education and training on the value of mHealth apps and knowledge of prescribable apps to be able to successfully change their prescribing habits to include apps. The future of sustainable and scalable app prescription requires a trustworthy electronic app repository of prescribable mHealth apps for GPs.

(KEYWORDS: mHealth apps; app prescription; general practice)

Introduction

More than 350,000 apps exist in the Medical and Health and Fitness categories in major app stores [1], with downloads and revenues in the billions [2]. Their popularity and potential to influence health-related behaviors make their integration to medical practice imminent [3]. Pragmatic studies of app prescription in primary care have been emerging around the world with varied interventions and results [4-6]. To assist the integration of apps into clinical practice, mobile health (mHealth) app repositories have been created, including the National Health Service App library in the United Kingdom [7], Health Navigator in New Zealand [8], and other private entities such as AppScript [9] and the Organization for the Review of Care and Health Applications [10].
Given the potential of mHealth apps to help improve the self-management of chronic conditions, we explored their value in general practice. Previously, in an overview of systematic reviews, we explored the possibility of simple integration of mHealth apps into the general practice setting and proposed a concept of “prescribable” mHealth apps. These were defined as proven effective (that is, shown to help achieve measurable clinical improvements in patients’ conditions), in addition to being standalone and currently available in the app stores [11].

We also explored the potential barriers to app integration in Australian general practice [12]. Patients expressed their preference for doctor-recommended apps; however, doctors were overwhelmed by the sheer number of available apps and faced 2 major barriers: not knowing of many prescribable apps and the lack of trustworthy source to access such apps. To address these barriers, we developed a brief implementation intervention. Objectives of this study were to explore the feasibility of app prescription by general practitioners (GPs) and to evaluate the effectiveness of an implementation intervention to increase uptake of app prescription.

Methods

Study Design and Setting

We employed a single-group, before-and-after design. Our study was conducted in the Australian general practice setting. Ethics approval was obtained from the Bond University Human Research Ethics Committee (#OB00017).

Participants and Recruitment

GPs currently working in Australia at least 2 days a week were eligible to participate in our study. Information about the study was distributed at 2 annual national GP conferences (GPDU2018 and GP18) and posted to a closed Facebook group called GPs DownUnder. Recruitment occurred from June 2018 through November 2018, and data collection occurred from September 2018 until May 2019. Upon completion of the study, GPs were thanked with Aus $50 gift cards.

Intervention

There were two parts to the intervention. First, prescription pads for 6 apps were developed (Figure 1). These apps were chosen because they address conditions relevant in general practice, have either direct trial evidence (This Way Up: Managing Depression, St. Vincent’s Hospital Sydney Ltd [13]; Tät – Pelvic floor exercises, Umeå University, Sweden [14]; Lose-It!, FitNow Inc [15,16]; CBT-i Coach, US Department of Veteran’s Affairs [17]) or indirect evidence from trials of similar apps (Smiling Mind, Smiling Mind Pty Ltd [18-20] and Quit Now: My QuitBuddy, Australian National Preventative Health Agency [21]). The apps also had to have stable content, were created or backed by trustworthy not-for-profit organizations, and were available for both Apple and Android phones. Five of the apps were freely available, and one (This Way Up: Managing Depression) had a one-time purchase price of Aus $59.99. The cost of apps was not an exclusion criterion as it will help assess if cost is a barrier to app prescription.

The app prescription pads had individually numbered pages with a tear-off design. Each app prescription page included the app’s full name and logo, download instructions, space for the patient’s name, the reason for prescription, and a disclaimer. Prescription pads were assembled onto an A4 display stand and mailed to participating GPs. A letter outlining the study timelines and procedures along with a short introduction to each app was included in the shipment.

The second part of the intervention was aimed at enhancing uptake. Short videos (2 minutes) demonstrating the content, functions, and features of the apps in detail were created for each app. A YouTube link to the video randomly selected for each participant was emailed following their second month’s reporting.

Our study aimed to change the prescribing behavior of GPs. Evidence suggests that behavioral interventions are more effective and sustainable when guided by behavior change techniques. Our prior research helped to identify the target behaviors [12]. We based our intervention on the Capability, Opportunity, Motivation, and Behavior model [22]. Capability to prescribe apps was addressed through the list of evidence-based apps and the introductory videos demonstrating the content, features, and function of the apps; opportunity was enabled through the purposefully designed stand with the prescription pads; and motivation was harnessed through the GPs’ expressed interest in the study that demonstrates their belief that app prescription would be a good thing to do [23].
Figure 1. The 6 app prescription pads, showing the front (A), with prescription details and script number in the bottom right corner, and back (B), with app download instructions and cost.

Procedures
At the beginning of the study, participants signed consent forms and answered the prestudy questionnaire via the web-based SurveyMonkey tool (SurveyMonkey Inc, San Mateo, CA). The survey collected demographic data, contact details, current app prescription rate in the preceding 2 weeks (self-reported in ranges: 0, 1-5, 6-10, >10 times), and level of confidence around app prescription.

The official commencement dates for the study were recorded as the date that each participant reported they started using the pads. Every 4 weeks following commencement, participants were asked to send a photo of the prescription pads electronically to the research team to provide details of the number of prescriptions for each app within that month. If participants took leave from work, the reporting dates were adjusted to allow for a full 4-week reporting period.
Qualitative semistructured interviews (10-15 minutes) were conducted and audio recorded at the end of the study, either face-to-face or by telephone, to gather feedback on the intervention. GPs were asked about their knowledge of other apps and relevant resources outside the study, including the Handbook of Non-Drug Interventions (HANDI) project by the Royal Australian College of General Practitioners, which includes a number of mHealth apps. Interviews were transcribed verbatim, coded by the lead researcher (OB), and thematically analyzed to determine the feasibility of the interventions, barriers, and solutions to the scalability of the intervention to Australian GPs. The thematic analysis was done in consultation with a second author (TH).

Sample Size
Prior data [12] indicated that the difference in the response before and after is normally distributed with a standard deviation of 10 and a baseline mean of 2 apps prescribed per month per GP. We calculated that we needed 24 participants to have 80% power (with $\alpha=.05$) to detect an increase of app prescription by at least 6 per month. Taking attrition into account, we planned to recruit 30 GPs for the study.

Outcomes
Data on app usage were collected for the 2-week period prior to study commencement and then every month for 4 months. The primary outcome of the study was the number of app prescriptions dispensed in total, as an average per month, per GP, and per GP per fortnight. We calculated the median number of apps recommended by a GP per fortnight using the following formula:

$$m = l + \frac{w(n/2-c)}{f}$$

where $l$ is the lower limit of the bin (range) containing the median, $w$ is the width of the bin, $n$ is the total population, $c$ is the cumulative count (frequency) up to $l$, and $f$ is the count in the median bin.

Prestudy raw numbers are provided in Table 1 ($m=1.7 \ [1 + (5(39/2-17))/19]$). Poststudy numbers are given in the Results section ($m=4.1 \ [1 + (5(39/2-0))/31]$).

Secondary outcomes were confidence around prescribing apps (measured on a 5-point Likert scale; prestudy and poststudy); the number of intervention video views and their impact on the subsequent prescription numbers; and attrition rate. In addition, the acceptability of the interventions to GPs and their feedback on the interventions were explored in semistructured interviews. Descriptive statistics were used to report the frequency of app use at each time point and confidence in app prescription. Qualitative data were analyzed thematically.

To conduct the overall analysis of the effect of video exposure on prescription rates, the 6 separate outcomes (1 for each app) were considered as one overall global outcome (individual monthly counts were not aggregated). Initially, a Poisson model was fitted with the (categorical) explanatory variables specified: the month (1 to 4), exposure to the video (yes/no), video (1 to 6), and interaction between exposure and video. To account for the 24 repeated measures collected for each GP (4 timepoints by 6 apps), a random intercept was fitted. Overdispersion was assessed using generalized chi-square divided by degrees of freedom. Due to evidence of overdispersion for the Poisson model (generalized chi-square/degrees of freedom=2.13), a negative binomial model was fitted and showed no evidence of over-dispersion (generalized chi-square/degrees of freedom=0.98).

Results
Overview
A total of 40 currently practicing Australian GPs were recruited for this study. One GP dropped out before the beginning of the study, and 3 GPs dropped out after the second and third data collection due to relocation and change of jobs. The full 4-month study was completed by 36 GPs, and we analyzed the data as intention-to-treat (ITT). The median age of the participants was 40 years, the median length of time in practice was 8.5 years, and participants worked a median of 4 days a week (Table 1).
Table 1. Participant demographics and prestudy characteristics, n=39.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>≤35</td>
<td>11 (28)</td>
</tr>
<tr>
<td>36-45</td>
<td>20 (51)</td>
</tr>
<tr>
<td>46-55</td>
<td>5 (13)</td>
</tr>
<tr>
<td>≥56</td>
<td>3 (8)</td>
</tr>
<tr>
<td><strong>Years in practice</strong></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>23 (59)</td>
</tr>
<tr>
<td>11-20</td>
<td>12 (31)</td>
</tr>
<tr>
<td>≥21</td>
<td>4 (10)</td>
</tr>
<tr>
<td><strong>Female gender</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>28 (72)</td>
</tr>
<tr>
<td><strong>Geographical distribution</strong></td>
<td></td>
</tr>
<tr>
<td>Queensland</td>
<td>21 (54)</td>
</tr>
<tr>
<td>New South Wales</td>
<td>9 (23)</td>
</tr>
<tr>
<td>Victoria</td>
<td>4 (10)</td>
</tr>
<tr>
<td>South Australia</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Western Australia</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Tasmania</td>
<td>1 (3)</td>
</tr>
<tr>
<td><strong>Days worked in a week</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8 (21)</td>
</tr>
<tr>
<td>3</td>
<td>9 (23)</td>
</tr>
<tr>
<td>4</td>
<td>13 (33)</td>
</tr>
<tr>
<td>≥5</td>
<td>9 (23)</td>
</tr>
<tr>
<td><strong>Number of apps prescribed in the 2 weeks prior to the study</strong></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>17 (44)</td>
</tr>
<tr>
<td>1-5</td>
<td>19 (49)</td>
</tr>
<tr>
<td>6-10</td>
<td>3 (8)</td>
</tr>
<tr>
<td><strong>Confidence level with app prescribing</strong></td>
<td></td>
</tr>
<tr>
<td>Not at all (1)</td>
<td>7 (18)</td>
</tr>
<tr>
<td>Not so (2)</td>
<td>12 (31)</td>
</tr>
<tr>
<td>Somewhat (3)</td>
<td>19 (49)</td>
</tr>
<tr>
<td>Very (4)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Extremely (5)</td>
<td>0</td>
</tr>
</tbody>
</table>

Prescriptions

In total, 1324 app prescriptions were dispensed over 4 months, with a mean of 331 prescriptions a month. Figure 2 illustrates the number of individual app prescriptions within the monthly totals. The number of apps prescribed per GP per fortnight increased from an imputed prestudy median of 1.7 to 4.1. Overall, the Smiling mind app was the most frequently prescribed (533/1324, 40%), followed by CBT-i Coach (242/1324, 18%), Managing Depression (167/1324, 13%), Lose-It! (155/1324, 12%), Quit Now: My QuitBuddy (134/1324, 10%), and Tät Pelvic floor exercises (93/1324, 7%).
Figure 2. Number of individual app prescriptions shown in the monthly totals.

Figure 3 illustrates the distribution of the total app prescription per GP. According to the ITT analysis, a median of 30 apps (range 6-111 apps) was prescribed per GP over the 4 months. Every GP prescribed at least one app per fortnight, 31 (31/39, 80%) GPs prescribed 1-5 apps, 7 (7/39, 18%) prescribed 6-10 apps, and 1 GP prescribed more than 11 apps. The GPs’ confidence around prescribing apps doubled from a mean of 2 (not so confident) before the study to 4 (very confident) at the end of the study: 0/39 not confident at all; 1/39 (3%) not so confident; 12/39 (31%) somewhat confident; 25/39 (64%) very confident; 1/39 (3%) extremely confident.

At the end of the study, the My QuitBuddy app video was viewed 8 times; the Smiling mind, Managing Depression, and Lose-It! app introduction videos were viewed 9 times each; the Tat-Pelvic floor exercise video was viewed 19 times; and the CBT-i Coach video was viewed 21 times. We were not able to track whether every GP watched the video sent to them. The effects of exposure to app videos are shown in Figure 4. Only two of the app videos had a significant effect on the subsequent app prescription numbers following the exposure to the video: Smiling Mind app prescription increased from 3-4 times per month to 6 times per month, and Lose-It! app prescription increased by one time. The full analysis is provided in Multimedia Appendix 1. A global test for the interaction between exposure and video showed strong evidence of heterogeneity (P<.001) indicating the treatment effects were different across the 6 apps. Therefore, we did not report an overall effect of the videos.
Figure 3. Distribution of total app prescription per general practitioner (GP). The red dot indicates the median (30 apps), the white dots indicate the participants who dropped out, and the dashed circle represents the participant who never commenced.
Qualitative Interviews
As per the ITT analysis, 39 GPs were interviewed at the conclusion of their participation in the study. Participants expressed their overall experience of prescribing apps as overwhelmingly positive. They liked the size of the prescription pad, the information included on it, ease of use, and integration into the workflow, with the most useful feature identified as the visual cue aspect of the stand. They also liked the short length of the videos, yet felt they contained sufficient details about the apps. Most GPs reported not downloading and interacting with the apps themselves. Although most reported having watched the allocated video, many did not recall the contents during the poststudy interviews.

Two of the 6 study apps were well known to the GPs: 28/39 GPs were already familiar with Smiling Mind, and 12/39 GPs were already familiar with Managing Depression. They had been recommending these apps to their patients even before the study and appreciated having a formal prescription to give out during the study. Among the other apps that GPs recommended, mindfulness and meditation apps (Calm, Headspace) were common. Mental health–related apps were the most frequently prescribed, and all GPs reported that the overall number of apps they prescribed is a reflection of the demographics of their patients and the prevalence of conditions encountered.

GPs reported that they might have prescribed the weight loss and pelvic floor exercise apps more frequently. Instead, they habitually referred patients to dietitians and physiotherapists or to programs and tools already compiled as the first line of intervention. None of the GPs, except for one, had watched, read, or received any other app-related content apart from the study intervention. Knowledge of HANDI was low, especially that apps were included in some HANDI entries. However, upon learning this, they all agreed that HANDI would be a reliable evidence-based app repository for GPs in Australia. The main barriers and facilitators to app prescription in general practice are shown in Table 2 along with illustrative quotes.
### Table 2. Key themes and illustrative quotes around barriers and facilitators of mobile health app prescription in general practice.

<table>
<thead>
<tr>
<th>Theme type, theme</th>
<th>Illustrative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Barriers</strong></td>
<td></td>
</tr>
<tr>
<td>Poor knowledge and familiarity of prescribable apps</td>
<td>“I think from a doctor, it's purely just knowledge of health apps.”</td>
</tr>
<tr>
<td></td>
<td>“From the GP's point of view, thinking about it, knowing which ones are good and which ones aren't.”</td>
</tr>
<tr>
<td></td>
<td>“challenging because I wasn't necessarily familiar myself with the details of the app in terms of using them myself or actually being able to really coach patients with using them. I guess that takes time to sit down and actually go through the apps.”</td>
</tr>
<tr>
<td>Prescribing habit</td>
<td>“Getting into the habit of having those things available was part of the process, trying to trigger the idea that I can do this was part of it.”</td>
</tr>
<tr>
<td></td>
<td>“I think trying to, in a busy consultation, trying to remember that as an option that we could recommend to people, because often you're so busy going, here, have this, do this, have this medication and then you often - adding some sort of self-help app into this is just part of getting more used to thinking about it as being an option.”</td>
</tr>
<tr>
<td>Cost of apps</td>
<td>“The depression one was quite an expensive app, that was quite prohibitive to a lot of people.”</td>
</tr>
<tr>
<td></td>
<td>“I guess I think cost definitely is a barrier for some patients, especially those that are in financial difficulty because they even ask for a referral to a bulk-billing psychologist.”</td>
</tr>
<tr>
<td>Patients’ capability and attitude towards mHealth apps</td>
<td>“I think they're probably for me the two big factors, is (one, the doctor's knowledge of them and) two, the patient perception of how important it is or the value of these health apps in terms of part of their management plan.”</td>
</tr>
<tr>
<td></td>
<td>“A lot of my main issue was the demographic of my patients. I didn't realize how much I would struggle to incorporate it because I actually have a huge percentage of elderly patients who don't even have smartphones and some of them that do, don't know how to use the apps properly.”</td>
</tr>
<tr>
<td>Consultation time</td>
<td>“time constraints, a lot of the time we're running behind and the app prescription is a slightly luxury, but when we have time and we're able to be thorough, of course, we can do it, but we don't always have that luxury of time”</td>
</tr>
<tr>
<td></td>
<td>“Time is just such a big issue because we're so time-pressured.”</td>
</tr>
<tr>
<td><strong>Facilitators</strong></td>
<td></td>
</tr>
<tr>
<td>Tailored education, face-to-face training, and information dissemination to increase knowledge of prescribable apps</td>
<td>“It's one message consistent and persistent. So if you've got a list that you're confident in, then why are you confident in it, what's the message behind and then you get it out as many ways as you can because none of us is looking at everything all the time. So if there's some way to get it out to the colleges, is there some way to get it out of the journals, is there somewhere to put it online somewhere that's an authoritative source, is there some way to get it out through the universities? Word of mouth is always good, influencers, social media…”</td>
</tr>
<tr>
<td></td>
<td>“Coming and meeting us and going through face to face, maybe demonstrating some, a bit like the drug reps do”</td>
</tr>
<tr>
<td></td>
<td>“I mean getting doctors early, so getting them through their training programs, getting them as GP registrars and making it part of there, I think that's where you're going to really get significant change.”</td>
</tr>
<tr>
<td>Meaningful familiarity with apps</td>
<td>“GP's own familiarity with the app, that if you're familiar with it, it's going to be much easier to prescribe than something that you have just head about or read about.”</td>
</tr>
<tr>
<td></td>
<td>“I think certainly the more hands-on you can get, I've done a couple or participated in a couple of webinars from the e-mental health stuff probably a year or two ago and that helped with my awareness of things, but my confidence I don't think improved too much, I think you've got to do them. You've either got to… Use it yourself or see it being used or at least be familiar with what it looks like.”</td>
</tr>
<tr>
<td>Trustworthy source of vetted prescribable apps</td>
<td>“I think having somebody external to narrow down the pool of apps and say this is a decent product, then you don’t mind recommending them in that way.”</td>
</tr>
<tr>
<td></td>
<td>“If it's coming from a reliable source like the university and say these are the apps we think are good quality apps to recommend, then I feel comfortable because there is so much information on the internet and app world, we don't know which is good quality and which is fake.”</td>
</tr>
<tr>
<td>Integration with existing software and workflow</td>
<td>“I think it would be brilliant to have an app that I could use for chronic disease management that actually was integrated, that the patients could potentially put data into that will then be integrated with my software, that would be fantastic.”</td>
</tr>
<tr>
<td></td>
<td>“Certainly, would help to have them integrated into our - the fact that we've prescribed them, into our software, medical software, so that we can just click a button to say recommended whichever app.”</td>
</tr>
</tbody>
</table>
This study demonstrated that it may be possible to increase the uptake of mHealth app prescription by providing an implementation intervention in the Australian general practice setting. The results demonstrated a total of 1324 app prescriptions by 39 GPs over 4 months and positive feedback from GPs about the intervention. The fortnightly number of apps prescribed per GP more than doubled compared to the prestudy level. However, identified barriers to app prescription uptake were poor knowledge of prescribable apps and insufficient familiarity with the apps to foster confident prescribing habits. Participants identified a need for a reliable prescribable app repository, preferably integrated with their electronic medical systems, and consistent and persistent messaging to increase the knowledge and familiarity of such apps.

The variation in the total individual tally of apps prescribed by participants may reflect differences in their personal digital propensity and flexibility in altering prescribing behavior. The reduction in the monthly app prescriptions after the first month could be related to the timing of the second and third reporting for about half of the participants. These occurred during the Christmas, New Year, and summer holidays in Australia, during which acute conditions dominate GP visits more than chronic conditions, which were the focus of the apps in the intervention.

The app explanation videos had varying effects on app prescription numbers. The results from the qualitative interviews showed that app prescription numbers were primarily dependent on the patient cohort and the prevalence of the conditions for which the intervention apps were intended. Thus, the short explanatory videos were informative but unlikely to be sufficient to influence complex behaviors such as prescribing. Perhaps, it would be more beneficial if video introduction and instructions for mHealth apps were developed for patients and given as part of the app prescription.

This is the first study to test the feasibility of an intervention to increase app prescription in Australian general practice. The overall attrition rate was low, and we analyzed the data as ITT, including those who dropped out of the study. Limitations include lack of access to electronic medical record data of the GP clinics to correlate the prevalence of conditions with the frequency of app prescription within the patient cohort. We aimed to recruit a sample of GPs representative of the national GP cohort; however, our participants’ median age of 40 years was younger than the national average of 50-55 years. Other limitations include a single-group pre-post study design, possible volunteer bias of the participants, and short time frame (4 months). Ideally, a randomized controlled trial should be conducted to establish the long-term effectiveness of the intervention with a large and representative sample for a longer duration. Due to the restrictions of available time and resources, we were unable to achieve this. Future studies should also opt for an electronic version of app prescription to improve sustainability and scalability. Another limitation is the analysis of qualitative data by a single researcher; however, the qualitative data result was a small part of our secondary outcome to primarily answer if the intervention was acceptable and feasible for practicing GPs.

There are few comparable studies of app recommendation in a primary care setting. A trial for an app prescription platform, AppSalut, in Spain involved 32 doctors who made 79 app recommendations in 5 months. Of the three apps they used, a medication adherence app was the most prescribed [4]. It sends the prescribed app to patients as text messages and can monitor and receive data on patients’ use and adherence to the system. In the United States, the Cambridge Health Alliance network of primary care clinics implemented a mental health app dissemination program, in which they evaluated mental health apps, selected 7 apps, and recommended these 7 apps in 12 primary care clinics [5]. Similar to the finding of our study, app prescriptions for anxiety and stress were the most frequently prescribed. An Australian study tested the feasibility of integrating mHealth apps into dietetic practice by asking 5 dietitians to use one chosen app for 12 weeks [6].

All of these studies provided training to the participating health care professionals to educate them about the study apps as well as the electronic systems they needed to use. The qualitative feedback from our participants also included the need for such

Illustrative quotes

<table>
<thead>
<tr>
<th>Theme type, theme</th>
<th>Illustrative quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual reminder or cue for prescribable apps</td>
<td>“having those pads in front of me made me think about it, the reminders and having a resource to go to.”</td>
</tr>
<tr>
<td>Patients’ capability and attitude towards mHealth apps</td>
<td>“most of the current population, the phone is the one thing that they carry around that they have with them all the time. Instead of - especially them being able to use it as an extra tool, they’re useful in the way of treating patients.”</td>
</tr>
<tr>
<td>Proof of benefits of apps as an alternative and or adjunct treatment</td>
<td>“sometimes the apps were very useful for patients who I was aware weren't able to afford other options. So for example, the pelvic floor exercises app would sometimes occur to me when I was talking to patients about the difficulties of accessing physiotherapy due to the cost and it would then prompt me to think, oh yes, actually I have an app that you could try at home without cost.”</td>
</tr>
<tr>
<td></td>
<td>“maybe some data showing that they are received well by patients, I guess. apps showing patient receptiveness and patient engagement”</td>
</tr>
</tbody>
</table>
training. However, because GPs often report being overworked, time-poor, and inundated with different information and offers, it would be challenging to organize out-of-hours training involving many GPs or train dedicated personnel to visit GP clinics during lunch hours, which was suggested by the GPs as a solution. The scalability of such an intervention would pose funding and logistical challenges.

One way to promote the sustainability and scalability of mHealth app integration into clinical practice is to provide an electronic repository of vetted and curated apps for health care professionals. In Australia, the Victoria Department of Health [24], Black Dog Institute [25], and HANDI project by the Royal Australian College of General Practitioners [26] offer small repositories of mHealth apps, but these organizations function under different jurisdictions with no national guideline in place. GPs in our study emphasized the need for a nationally accessible repository of a select few prescribable apps that are relevant to general practice that is safe, reliable, and easy to navigate.

We found that mHealth app prescription is feasible in a general practice setting in Australia by addressing previously identified practical barriers to mHealth app prescription. Our implementation intervention was effective in increasing app prescription. However, the future of app prescription depends on efforts to increase GPs’ knowledge of prescribable apps as well as a dedicated trustworthy app repository for GPs.

Acknowledgments
We would like to thank Professor Mark Jones for statistical assistance.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Statistical data for intervention video impact.
[DOCX File, 49 KB - mhealth_v8i6e16497_app1.docx ]

Multimedia Appendix 2
App prescription pads being used.
[MMOV File, 3569 KB - mhealth_v8i6e16497_app2.MOV ]

References


Abbreviations

GP: general practitioner.
HANDI: Handbook of Non-Drug Interventions.
ITT: intention to treat.
mHealth: mobile health.

Edited by G Eysenbach; submitted 04.10.19; peer-reviewed by J Chen, A Vallée, J LandeI, L Muldon; comments to author 23.12.19; revised version received 21.02.20; accepted 26.03.20; published 01.06.20.

Please cite as:
Byambasuren O, Beller E, Hoffmann T, Glasziou P
mHealth App Prescription in Australian General Practice: Pre-Post Study
JMIR Mhealth Uhealth 2020;8(6):e16497
URL: https://mhealth.jmir.org/2020/6/e16497
doi:10.2196/mhealth.4328
PMID:31304297
Assessing the Impact of Patient-Facing Mobile Health Technology on Patient Outcomes: Retrospective Observational Cohort Study

Courtenay R Bruce, MA, JD; Patricia Harrison, BSN, MBA, RN; Tariq Nisar, MPH; Charlie Giammattei, BSc; Neema M Tan, MHA; Caitlin Bliven, BA; Jamie Shallcross, BSc; Aroub Khleif, PhD; Nhan Tran, BSc; Sayali Kelkar, MPH; Noreen Tobias, BSc; Ana E Chavez, BSN, RN; Dana Rivera, BSc; Angela Leong, BSN, RN; Angela Romano, DPM, MBA; Josh R Sol, BSc; Kayla Gutierrez, MSN, RN, APRN; Christopher Rappel, MSN, RN; Eric Haas, MD; Feibi Zheng, MD, MBA; Kwan J Park, MD; Stephen Jones, MD, MSHI; Paul Barach, MD, MPH; Roberta Schwartz, PhD

System Quality & Patient Safety, Houston Methodist Hospital System, Houston, TX, United States
Center for Outcomes Research, Houston Methodist Research Institute, Houston, TX, United States
CareSense, Philadelphia, PA, United States
Information Technology Department, Houston Methodist Hospital, Houston, TX, United States
Information Technology Portfolio Management Office, Houston, TX, United States
Cardiovascular Surgery Associates, Houston Methodist Hospital, Houston, TX, United States
Houston Methodist Orthopedic and Sports Medicine - The Woodlands, Houston Methodist The Woodlands, The Woodlands, TX, United States
Clinical Design and Planning, Houston Methodist Hospital System, Houston, TX, United States
Houston Methodist Specialty Physician Group, Houston Methodist Orthopedics & Sports Medicine, Houston Methodist Hospital, Houston, TX, United States
Houston Methodist Orthopaedics, Houston Methodist Hospital, Houston, TX, United States
Houston Methodist Specialty Physician Group, Houston Methodist Orthopedics & Sports Medicine, Houston Methodist Hospital, Houston, TX, United States
Houston Methodist Specialty Physician Group, Houston Methodist Orthopedics & Sports Medicine, Houston Methodist Hospital, Houston, TX, United States
Department of Surgery, Houston Methodist Hospital, Houston, TX, United States
Jefferson College of Population Health, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Courtenay R Bruce, MA, JD
System Quality & Patient Safety
Houston Methodist Hospital System
6565 Fannin Street
Houston, TX, 77030
United States
Phone: 1 2816209040
Email: crbruce@houstonmethodist.org

Abstract

Background: Despite the growth of and media hype about mobile health (mHealth), there is a paucity of literature supporting the effectiveness of widespread implementation of mHealth technologies.

Objective: This study aimed to assess whether an innovative mHealth technology system with several overlapping purposes can impact (1) clinical outcomes (ie, readmission rates, revisit rates, and length of stay) and (2) patient-centered care outcomes (ie, patient engagement, patient experience, and patient satisfaction).

Methods: We compared all patients (2059 patients) of participating orthopedic surgeons using mHealth technology with all patients of nonparticipating orthopedic surgeons (2554 patients). The analyses included Wilcoxon rank-sum tests, Kruskal-Wallis tests for continuous variables, and chi-square tests for categorical variables. Logistic regression models were performed on categorical outcomes and a gamma-distributed model for continuous variables. All models were adjusted for patient demographics and comorbidities.

Results: The inpatient readmission rates for the nonparticipating group when compared with the participating group were higher and demonstrated higher odds ratios (ORs) for 30-day inpatient readmissions (nonparticipating group 106/2636, 4.02% and...
participating group 54/2048, 2.64%; OR 1.48, 95% CI 1.03 to 2.13; \( P = .04 \), 60-day inpatient readmissions (nonparticipating group 194/2636, 7.36% and participating group 85/2048, 4.15%; OR 1.79, 95% CI 1.32 to 2.39; \( P < .001 \)), and 90-day inpatient readmissions (nonparticipating group 261/2636, 9.90% and participating group 115/2048, 5.62%; OR 1.81, 95% CI 1.40 to 2.34; \( P < .001 \)). The length of stay for the nonparticipating cohort was longer at 1.90 days, whereas the length of stay for the participating cohort was 1.50 days (mean 1.87, SD 2 vs mean 1.50, SD 1.37; \( P < .001 \)). Patients treated by participating surgeons received and read text messages using mHealth 83% of the time and read emails 84% of the time. Patients responded to 60% of the text messages and 55% of the email surveys. Patients were least responsive to digital monitoring questions when the hospital asked them to do something, and they were most engaged with emails that did not require action, including informational content. A total of 96% (558/580) of patients indicated high satisfaction with using mHealth technology to support their care. Only 0.40% (75/2059) patients opted-out of the mHealth technology program after enrollment.

**Conclusions:** A novel, multicomponent, pathway-driven, patient-facing mHealth technology can positively impact patient outcomes and patient-reported experiences. These technologies can empower patients to play a more active and meaningful role in improving their outcomes. There is a deep need, however, for a better understanding of the interactions between patients, technology, and health care providers. Future research is needed to (1) help identify, address, and improve technology usability and effectiveness; (2) understand patient and provider attributes that support adoption, uptake, and sustainability; and (3) understand the factors that contribute to barriers of technology adoption and how best to overcome them.

(JMIR Mhealth Uhealth 2020;8(6):e19333) doi:10.2196/19333

**KEYWORDS**

mHealth; patient-centered care; patient satisfaction; length of stay; patient activation; patient empowerment; patient engagement; patient involvement; hospital stay; communication programs

**Introduction**

**Background**

Improving patient safety and quality of care remains to be the key goal of health care systems. In 2001, the Institute of Medicine identified patient-centered care as a health care quality indicator [1,2]. Within the past decade, top health care organizations have embraced mobile health (mHealth) as part of their patient-centered initiatives and their drive to achieve the quadruple aim [3-7]. Indeed, 18 of the 20 hospitals listed on the 2019-2020 Honor Roll for the US News and World Report have adopted at least one patient-centered mHealth technology [3]. Funding for mHealth technologies exceeded US $3 billion in the United States in the first 6 months of 2019 [4], and the European Union has committed to investing US $24 billion in mHealth technologies [5,6]. In spring 2020, coronavirus disease 2019 (COVID-19) resulted in the rapid global uptake of patient-facing mHealth technologies to support patient needs and protect health care providers while promoting social distancing [7].

Despite this growth, there remains a paucity of data regarding whether mHealth technologies have scientific merit and are effective to warrant such widespread implementation [8]. Specifically, it is unclear whether mHealth technologies positively impact clinical outcomes, such as hospital length of stay, patient readmissions, or complications [9]. It is also unclear whether mHealth technologies impact patient-centered care outcomes, defined and measured by the perceptions of patient experiences, patient engagement, or patient satisfaction [10,11]. Existing studies focus on telemonitoring mHealth technologies and clinical outcomes, often showing unremarkable results. For instance, one large-scale study showed no reduction in readmission rates [12], and a few single-institution assessments demonstrated that telemonitoring using mHealth technologies could modestly improve patient adherence [13,14]. mHealth technologies beyond telemonitoring, such as tele-education, teleconsultation, and digital navigation, are largely unexplored and have not shown meaningful impacts on patient outcomes [5-14].

**Objectives**

We sought to assess whether an innovative mHealth technology system with several overlapping purposes, specifically tele-education and telemonitoring features, can impact (1) clinical outcomes (ie, readmission rates, revisit rates, and length of stay) and (2) patient-centered care outcomes (ie, patient engagement, patient experience, and patient satisfaction).

**Methods**

**Setting and Context**

This study was a retrospective, observational cohort study. We retrospectively analyzed all patients treated by orthopedic surgeons in our hospital system who actively participated in using mHealth technology from January 1 to December 31, 2019, and compared them with all patients of nonparticipating orthopedic surgeons during the same period. The implementation phase was staggered in a phased rollout. Surgeons who were not yet approached did not participate, and the technology was not offered to their patients. All patients in both groups underwent a primary total joint (hip or knee) replacement (TJR). This observational cohort study was approved by the hospital system’s institutional review board. The hospital system consists of one 2264-bed tertiary academic medical center located in Houston, Texas, along with 7 community hospitals (300-700 beds) in the suburbs of Houston, Texas.

We excluded patients from the analysis who indicated any language other than English as their preferred language (percentages of which are reported in the Results section) to mitigate selection bias. We compared patients who only

http://mhealth.jmir.org/2020/6/e19333/
preferred the English language and used mHealth technology with all patients who preferred the English language and used mHealth technology.

We also conducted an uptake analysis of the first few months in which mHealth technology was available (January 1, 2019-April 30, 2019). Our goal in conducting a subset analysis was to show a robust biological gradient [15]. The presence of a dose-response relationship supports the causal association between an exposure and an effect and durability of the results, and, by extension, we sought to show the external generalizability benefits to the overall population [15,16].

The Participating Cohort
The technology consisted of a digital education and monitoring platform, CareSense, (MedTrak, Inc) for patients using their computers or mobile devices with text and email messages in English about their medical condition. Text messages were automatically delivered over the cellular network (SMS texts) via an internet connection. Patients did not need to download an mHealth app or go to a web-based patient portal for emails and texts to be transmitted. Patients did have the option, however, to access a secure, web-based portal to receive all messages in one area.

The text messages were converted to automated phone calls in cases where patients did not have text messaging capabilities. One or two messages were sent each day for the 20 days before surgery, were stopped when the patient was admitted to the hospital, and resumed once or twice a day for 30 days following the patient’s hospital discharge. We refer in this paper to the full sequence of messages throughout the 50-day period as the pathway.

Mobile Health Technology Content
The content of the pathway was primarily designed to achieve several purposes; (1) provide education, (2) monitor health and recovery, (3) provide key reminders to needed actions or taking of medication, and (4) ensure resolution of patients’ action items (Table 1).

Table 1. Example of patient messages and clinical domains.

<table>
<thead>
<tr>
<th>Period the message was sent and purpose of the message</th>
<th>Example messages</th>
<th>Clinical domains or goals</th>
<th>Response rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presurgery Educating Questions to ask your physician about preparing your home for surgery</td>
<td>Promoting preparation and understanding</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Presurgery Closing action items “Have you scheduled your visit for presurgery lab work?” and “Have you completed your presurgical clearance paperwork?”</td>
<td>Promoting adherence and self-management</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>Presurgery Monitoring “Have you checked your hemoglobin A1C? Was it over a 7?”</td>
<td>Optimizing health in preparation for surgery</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Closing action items “Do you have someone who can pick you up after surgery, regardless of what hour you are discharged?”</td>
<td>Discharge planning</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Closing action items “Do you have someone available after surgery who can help you for the first 24-48 hours?”</td>
<td>Enlisting help of social supports</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Educating Tips on how to take care of your wound when to call our office ASAP</td>
<td>Monitoring and managing symptoms</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Educating Tips on mobility exercises</td>
<td>Optimizing health following surgery</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Closing action items “Have you scheduled your follow-up appointment yet?”</td>
<td>Outpatient follow-up</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Postsurgery Monitoring “Take a look at your incision site. Is it sore, very red, puffy…?”</td>
<td>Monitoring and managing symptoms</td>
<td>89</td>
<td></td>
</tr>
</tbody>
</table>

aASAP: as soon as possible.

Generally, time-sensitive and short messages were sent via text messaging, typically consisting of alphabetic and numeric characters, and longer educational messages were sent via email. Table 1 demonstrates examples of messages for all 4 purposes.

The messages were unidirectional or bidirectional. For unidirectional messages, there was no expectation of patient response. Unidirectional messages were educational or informative in nature. The bidirectional messages were sent to solicit patient responses using close-ended questions (Table 1). The bidirectional messages allowed clinicians to monitor patients’ health and recovery or, alternatively, to ensure that the patient completed important action items before and/or following their surgery.

When patients responded to bidirectional messages in a concerning way, an alert was automatically generated and routed to their health care professionals. For example, one bidirectional message read, “Please identify your pain level. Press 1 for no pain or mild pain (1-3 on a pain scale); Press 2 for moderate pain (4-6 on a pain scale); Press 3 for severe pain (7-8 on a pain scale); Press 4 for extreme pain (9-10 on a pain scale).” A patient who responded by pressing the numbers 3 or 4 generated an alert that was sent to the health care team (ie, medical assistants...
[MAs or nurses] via email, letting them know that a patient had responded to a monitoring question that required their timely response.

**Procedures**

During the surgery scheduling process, hospital schedulers of participating surgeons asked all English-speaking patients undergoing TJR whether they would be willing to receive digital messages. If so, the scheduler used the electronic medical record to activate the pathway. Clinicians and schedulers did not need to push messages to patients, as all messages were transmitted automatically via text/phone calls and email once the scheduler activated the pathway.

Patients could opt-out at any time after enrollment by typing the word stop in response to any text messages to stop receiving email and text messages. For patients who did not have text messaging abilities, they could press a number on their telephone number pad in response to phone calls to stop all phone calls and email messages.

The education provided to patients was minimal, largely because they did not need to download an app or go to a portal to receive messages. The messages were transmitted automatically for patients who agreed to receive them, requiring little technical expertise on the part of the patients. Patients were directed to call one hospital-based employee who had content and technical expertise with any questions about mHealth technology.

**Outcome Measures**

**Readmissions and Revisit Rates**

Hospital readmissions were defined as any subsequent unplanned inpatient admission to any of our system-based acute care facilities occurring within 30, 60, and 90 days of hospital discharge following the qualifying total joint operations. Only unplanned inpatient admissions (for any cause) to short-term acute care, excluding transfer encounters, qualified as readmission for the study, as our inclusion and exclusion criteria for calculation were consistent with the Centers for Medicaid and Medicare Services specifications and methodological standards [17].

The patient revisit rates were defined as any visit to an acute care facility that occurred within 30, 60, and 90 days after discharge following a qualifying total joint operation, aside from unplanned inpatient admissions—namely, emergency department visits; unplanned, unscheduled outpatient visits; and observation status visits [17].

The hospital readmissions and revisit rates were calculated based on reviewing electronic medical records for all elective primary total knee or total hip replacement operations performed between January 1 and December 31, 2019. Patients who were readmitted to the same hospital on the same calendar day of discharge for the same diagnosis as the index admission were considered to have 1 single continuous admission (ie, 1 index admission), and patients who were readmitted for a different condition from the index admission were considered to have a readmission within the measure. The analysis excluded staged surgical procedures by identifying the patients who were readmitted for another primary hip or knee procedure within the 30-, 60-, or 90-day periods.

**Length of Stay**

The hospital length of stay comprised the entire length of hospitalization and was calculated using the admission date until the discharge date.

**Other Prespecified End Points**

Patient-centered care outcomes included patient engagement, patient experience, and patient satisfaction. Patient engagement was defined as the degree to which the patient engaged with the mHealth technology [18,19]: (1) minimal engagement (ie, the patient read ≤25% of all messages and responded to ≤25% of all messages in the pathway, as indicated by a read receipt), (2) moderate engagement (ie, the patient read 26%-50% of all messages and responded to 26%-50% of all messages), or (3) high engagement (ie, the patient read ≥51% of all messages and responded to ≥51% of all messages). The patient had to read ≥51% of all messages and respond to ≥51% of all messages to be considered highly engaged. If they read messages but did not respond at a level of ≥51%, they would not meet the engagement threshold. We chose these thresholds consistent with the literature on patient engagement, where researchers proposed that empirical thresholds of engagement with mHealth technology must be met to show sufficient engagement [8,19].

Patient experience was defined as any process observable by patients, including their subjective experiences (eg, quality of communication) and their objective experiences (eg, how often communication occurred) [20]. We analyzed patient experiences by evaluating the patients’ responses to the validated, reliable Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey [21], regarding 4 questions:

1. Before giving you any new medicine, how often did the hospital staff describe possible side effects in a way you could understand? (4-point Likert scale, never to always)
2. How often did the hospital staff tell you what the medicine was for? (4-point Likert scale, never to always)
3. Did doctors, nurses, or other hospital staff talk with you about whether you would have the help you needed when you left the hospital? (yes or no)
4. Did you get information in writing about what symptoms or health problems to look out for after you left the hospital? (yes or no)

We compared the experience scores for each of these 4 questions for patients in the participating group as compared with those in the nonparticipating group.

Patient satisfaction was defined as whether patients’ expectations were met [10]. For the participating cohort, we analyzed the patients’ collective responses (Likert scale 1-5: strongly agree to strongly disagree) to a text question that we embedded in the pathway which asked, “How much do you agree with the following statement: It was helpful for me to receive reminders and emails from this program.” We chose this question because questions on helpfulness and ease of use are considered the most frequently used questions in most validated instruments for patient satisfaction [9,19,20].
Data Analysis

We used the two-tailed \( t \) test and Wilcoxon test for continuous variables and chi-square test for categorical variables to compare the baseline characteristics between the 2 groups [16]. The data about patient demographics were available for 100% of patients and providers. The clinical end points were compared using chi-square tests. Logistic regression models using the generalized estimated equation (GEE) method accounting for repeated measurements were performed on the categorical outcomes and the gamma-distributed model of the continuous variable. All tests for significance were two-tailed, using an alpha level of .05, and 95% CIs were provided. All statistical analyses were performed using R version 3.6.0 [21]. A multivariate regression analysis was conducted to minimize confounding factors and their potential impacts, such as age, gender, and comorbidities.

Results

Diversity of the Sample

Patients

A total of 2059 patients who underwent TJR were treated by participating surgeons using mHealth technology from January to December 2019, whereas 2554 patients who underwent TJR were treated by nonparticipating surgeons who did not use mHealth technology during the same period. To minimize the possibility of language being a confounding variable in the study, 162 patients were excluded from the analysis, including 119 patients who were not enrolled with mHealth technology who indicated a preference for a language other than English, and 43 patients who were enrolled in the mHealth technology who indicated a preference for a language other than English. This resulted in 2059 patients who underwent TJR treated by participating surgeons and 2554 patients who underwent TJR treated by nonparticipating surgeons. Only 0.30% (59/19,667) patients declined to enroll in the mHealth technology program.

Separately, we also conducted an analysis of the first few months when patients who underwent TJR treated by participating surgeons were offered mHealth technology (January 1, 2019-April 30, 2019) as compared with patients of nonparticipating surgeons. These results, which were similar to the January 1 to December 31, 2019, outcomes on all measures, can be found in Multimedia Appendix 1.

Demographics

The differences in the baseline characteristics of the patients are described in Table 2. Overall, the mean age of the patients was 68 years for the nonparticipating group and 67 years for the participating groups (\( P<.05 \)). There were 57.75% (1475/2554) of females in the nonparticipating group and 60.42% (1244/2059) in the participating group (\( P=.07 \)). The preferred language, ethnicity, and race were nearly identical in the nonparticipating and participating cohorts (Table 2).

<table>
<thead>
<tr>
<th>Variables</th>
<th>Nonparticipating cohort (n=2554)</th>
<th>Participating cohort (n=2059)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>68.07 (9.76)</td>
<td>66.83 (9.63)</td>
<td>&lt;.05 a</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>.07</td>
</tr>
<tr>
<td>Female</td>
<td>1475 (57.75)</td>
<td>1244 (60.42)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1079 (42.25)</td>
<td>815 (39.58)</td>
<td></td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>.83</td>
</tr>
<tr>
<td>White</td>
<td>2100 (82.22)</td>
<td>1687 (61.93)</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>332 (13.00)</td>
<td>263 (12.77)</td>
<td>.85</td>
</tr>
<tr>
<td>Asian</td>
<td>54 (2.11)</td>
<td>49 (2.38)</td>
<td>.61</td>
</tr>
<tr>
<td>Native American</td>
<td>6 (0.23)</td>
<td>0 (0.00)</td>
<td>.50</td>
</tr>
<tr>
<td>Others</td>
<td>28 (1.10)</td>
<td>14 (0.68)</td>
<td>.19</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td>.63</td>
</tr>
<tr>
<td>Hispanics</td>
<td>175 (6.85)</td>
<td>133 (6.46)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanics</td>
<td>2360 (92.40)</td>
<td>1913 (92.91)</td>
<td>.55</td>
</tr>
<tr>
<td>Declined to answer</td>
<td>19 (0.74)</td>
<td>1297 (0.63)</td>
<td>.78</td>
</tr>
<tr>
<td>Comorbidities, n (%)</td>
<td></td>
<td></td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Hypertension</td>
<td>829 (32.46)</td>
<td>792 (38.47)</td>
<td></td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>578 (22.63)</td>
<td>566 (27.49)</td>
<td>&lt;.0</td>
</tr>
<tr>
<td>Unilateral primary osteoarthritis</td>
<td>1112 (43.54)</td>
<td>1205 (58.52)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Gastroesophageal reflux disease</td>
<td>445 (17.42)</td>
<td>4410 (21.42)</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

Table 2. Sample patient characteristics of nonparticipating and participating groups.

aThe italicized values are statistically significant at a level of .05.
The surgeon demographics were similar between the nonparticipating and participating cohorts, and the differences were almost negligible, except for 1 fundamental attribute: the nonparticipating surgeons had higher baseline (ie, pre-mHealth technology) patient satisfaction scores compared with the participating surgeons (22/22, 100% compared with 92% baseline patient satisfaction scores, respectively; \( P < .05 \); Table 3).

### Table 3. Comparison between participating surgeons and nonparticipating surgeons.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participating surgeons (n=22)</th>
<th>Nonparticipating surgeons (n=25)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experience (years), mean (SD)</td>
<td>19.73 (12.53)</td>
<td>22.56 (9.23)</td>
<td>.39</td>
</tr>
<tr>
<td>Case volume, median (IQR)</td>
<td>166.5 (66.75-258.5)</td>
<td>107 (18.5-181.25)</td>
<td>.13</td>
</tr>
<tr>
<td>Readmission, median (IQR)</td>
<td>2.02 (1.65-2.31)</td>
<td>2.10 (1.10-3.1)</td>
<td>.97</td>
</tr>
<tr>
<td>Patient satisfaction scores, median (IQR)</td>
<td>92 (88.25-96)</td>
<td>100 (92-100)</td>
<td>&lt;.05 ( ^{a} )</td>
</tr>
</tbody>
</table>

\( ^{a} \)The italicized values are statistically significant at a level of .05.

### Patient Comorbidities

The comorbidities differed in both groups (Table 2), with the participating group having more comorbidities at baseline (ie, before enrolling them in the pathway). Specifically, 32.46% (829/2554) of the patients in the nonparticipating group were treated for hypertension, with 38.47% (792/2059) of patients in the participating group treated for the same condition (\( P < .05 \)). The incidence of hyperlipidemia occurred at a higher rate among patients in the participating group than in the nonparticipating group (566/2059, 27.49% vs 578/2554, 22.63%, respectively; \( P < .05 \)). Other comorbidities included gastroesophageal reflux diseases, unilateral primary osteoarthritis of the right hip, and unilateral primary osteoarthritis left knee, which also impacted patients within the participating group at higher rates than those in the nonparticipating group (4410/2059, 21.42% participating vs 445/2554, 17.42% nonparticipating for reflux; \( P < .05 \) and 1205/2059, 58.52% participating vs 1112/2554, 43.54% nonparticipating for osteoarthritis; \( P = .03 \)).

### Outcome Measures

### Readmissions and Revisit Rates

Within 30 days of surgery, 106/2636, 4.02% of inpatient readmissions occurred for the nonparticipating group, and 54/2048, 2.64% of inpatient readmissions occurred for the participating group (\( P = .001 \); Table 4).

Within 60 days after surgery, 194/2636, 7.36% inpatient readmissions occurred in the nonparticipating group, and 85/2048, 4.15% inpatient readmissions occurred in the participating group (\( P < .001 \)). Within 90 days after surgery, 261/2636, 10.00% of inpatient readmissions occurred in the nonparticipating group, and 115/2048, 5.62% of inpatient readmissions occurred in the participating group (\( P < .05 \)). After adjusting for demographics and comorbidities (hypertension, hyperlipidemia, gastroesophageal disease, and osteoarthritis), a multivariate logistic model (GEE) demonstrated that the nonparticipating group had a higher odds ratio (OR) for 30-day inpatient readmissions (OR 1.48, 95% CI 1.03 to 2.13; \( P = .04 \)), 60-day inpatient readmissions (OR 1.79, 95% CI 1.32 to 2.39; \( P < .001 \)), and 90-day inpatient readmission rates (OR 1.81, 95% CI 1.40 to 2.34; \( P < .001 \)) when compared with the participating group (Figure 1). The patient revisit rates are shown in Table 4.
Table 4. Hospital readmissions and revisit rates analyses.

<table>
<thead>
<tr>
<th>Outcome variablea,b</th>
<th>Control</th>
<th>Intervention</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Value, n (%)</td>
<td>N</td>
</tr>
<tr>
<td>Hospital readmission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 days</td>
<td>2636</td>
<td>106 (4.02)</td>
<td>2048</td>
</tr>
<tr>
<td>60 days</td>
<td>2636</td>
<td>194 (7.36)</td>
<td>2048</td>
</tr>
<tr>
<td>90 days</td>
<td>2636</td>
<td>261 (9.90)</td>
<td>2048</td>
</tr>
<tr>
<td>Emergency department visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 days</td>
<td>2594</td>
<td>134 (5.17)</td>
<td>2081</td>
</tr>
<tr>
<td>60 days</td>
<td>2594</td>
<td>183 (7.05)</td>
<td>2081</td>
</tr>
<tr>
<td>90 days</td>
<td>2594</td>
<td>206 (7.94)</td>
<td>2081</td>
</tr>
<tr>
<td>Unplanned, unscheduled outpatient visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 days</td>
<td>2571</td>
<td>24 (0.93)</td>
<td>2064</td>
</tr>
<tr>
<td>60 days</td>
<td>2571</td>
<td>81 (3.15)</td>
<td>2064</td>
</tr>
<tr>
<td>90 days</td>
<td>2571</td>
<td>129 (5.02)</td>
<td>2064</td>
</tr>
<tr>
<td>Observation status visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 days</td>
<td>2556</td>
<td>60 (2.35)</td>
<td>2059</td>
</tr>
<tr>
<td>60 days</td>
<td>2556</td>
<td>95 (3.72)</td>
<td>2059</td>
</tr>
<tr>
<td>90 days</td>
<td>2556</td>
<td>125 (4.89)</td>
<td>2059</td>
</tr>
</tbody>
</table>

aHospital readmissions were defined as “any subsequent unplanned inpatient admission to any acute care facility which occurred within 30, 60, and 90 days of discharge following the qualifying total joint operations.”
bRevisit rates were defined as “any visit to an acute care facility which occurred within 30, 60, and 90 days of discharge following qualifying total joint operations besides unplanned inpatient admissions—namely, emergency department visits, unplanned, unscheduled outpatient visits, and observation status visits” [11].
cThe italicized values are statistically significant at a level of .05.

Figure 1. Odds ratio for participating versus nonparticipating patients. The bars represent 95% CIs and the dots represent the odds ratio.

Length of Stay
The average length of stay for inpatient hospitalization for the nonparticipating group was 1.87 days as compared with 1.50 days for the participating group (P<.001).

Patient Engagement
There were 39 unidirectional text messages over 50 days. Patients read their text messages 90% of the time (median), indicated by a read receipt. Of the 39 messages, patients tended
to be most inclined to read text messages that included tips on hip care precautions, information on when to stop medications, and what to do the night before surgery. The text message patients were least inclined to read included information on getting presurgery clearance, smoking cessation, and preadmission testing (Table 1).

There were 29 unidirectional emails over the 50-day period. Patients read emails 89% of the time. The email messages patients read the most were getting to know your health care team, planning for your return home, commonly asked questions and answers, and pain medication details. The emails that patients read the least involved information about what surgery entails (videos) and infection and blood clot prevention strategies.

There were 13 bidirectional text messages over the 50-day pathway. Patients responded to messages 54% (median) of the time. Patients were most inclined to respond to questions about their mobilization and incision sites. They were least likely to respond to questions about whether they obtained presurgical clearance, registered for preoperative educational classes, or if they scheduled an appointment with preadmission testing.

MAs and nurses were alerted whenever a patient responded to a bidirectional question in a concerning way, typically because a patient responded to a question on pain levels in a way that indicated severe pain or, alternatively, the patient indicated unusual redness or excessive bleeding at the incision site. The inbound messages to MAs and nurses were minimal, suggesting that patients rarely responded to a bidirectional question in a concerning way. Each office received an average of 2 notifications each week for all of their patients.

**Patient Experience**

With regard to the medication questions on the HCAHPS survey, there were 428 responses to medication questions in the nonparticipating group, with an average of 56% of patients reporting that hospital staff *always* discussed side effects in a way that patients could understand, and an average of 86% of patients reported that the hospital staff *always* described the purpose of new medications (P values can be found in Table 5).

There were 485 responses to the same question from the participating group, with an average of 58% of patients reporting that the hospital staff *always* discussed medication side effects in a way patients could understand, and an average of 84% of patients reported that hospital staff *always* described the purpose of new medications.

The composite score for the medication questions was 71.3% for the participating cohort and 71% for the nonparticipating cohort.

**Patient Satisfaction With the Mobile Health Technology**

A total of 75/2059 (0.4%) patients opted-out of the mHealth technology program after enrollment by texting *stop* in response to text messages, all of whom opted-out in the postdischarge window, and usually did so toward the end of the pathway. We asked patients using mHealth technology whether they found it helpful to receive text messages, reminders, and emails by way of the mHealth technology from their health care providers. A total of 580 of the 2059 patients (28%) responded to the satisfaction question; 390 (67%) responded that they strongly agree that the participation was helpful; 166 patients (29%) responded that they agree that the mHealth technology was helpful; 8 patients (1.7%) were undecided; and 6 patients (1.4%) disagreed that mHealth technology was helpful. Only 6 patients (1%) strongly disagreed.

**Discussion**

**Principal Findings**

We found that patient-facing mHealth technology may positively impact patient readmissions and length of stay in significant ways, even after a few months of implementation of the technology. Patients cared for by participating surgeons were readmitted at a significantly lower rate as compared with patients of nonparticipating surgeons within 30, 60, and 90 days after surgery in both the subset analysis (Multimedia Appendix 1) and in the full analysis presented in the Results section (January 1-December 31, 2019). Furthermore, the length of hospital stay for patients cared for by the participating surgeons was about one-third less than that for patients of nonparticipating surgeons in both the subset and full analyses. We also found significant improvements in patient-centered care measures in the subset and full analyses.

The importance and significance of these findings are remarkable given that mHealth technologies are considered a way of the future, and have been shown to be remarkably...
popular and essential during the COVID-19 pandemic [4,22-24]. Our findings run counter to the notion that after technology adoption, the reduction of readmission rate will necessarily take months or years to achieve or that they cannot be sustained over time [25,26].

Of note, we found differences in the comorbidities, suggesting that patients cared for by participating surgeons were sicker than patients cared for by nonparticipating surgeons, at least at the time of enrollment in the mHealth technology. This finding is interesting in that if there were patient selection biases present, one might expect that health care professionals would favor enrolling healthier patients on mHealth technology [16,27-29]. Unfortunately, the opt-out rate for patients cared for by participating surgeons was too small of a cohort to discern whether there were any differences between the opt-in versus opt-out patient cohorts in the participating surgeon group. It is possible that the sicker patients of the participating surgeons were self-selecting to participate, which can be validated by future research.

We also found that mHealth technologies can impact multiple outcomes at the same time, including clinical outcomes and patient-centered care measures. These findings lend support to the notion that patient-centered care processes likely enhance several quality dimensions simultaneously [2]. Specifically, although existing empirical evidence suggests there may only be a modest positive association between patient-centered care and clinical outcomes [11,12,18,19,30], there is still a strong reason to use a systems-based approach to improve multiple aspects of care quality. Quality improvement efforts aimed at enhancing patient-centered care might improve infrastructure and processes, resulting in broader quality of care improvements on multiple levels [20].

The broad experience of successful use of mHealth technologies during the COVID-19 pandemic shows that effective technologies go beyond merely monitoring purposes by integrating the informational and educational needs of the patient/family/caregivers. Indeed, we theorize that the primary reason for the shorter length of stay for patients in the participating cohort is by virtue of the educational benefits of mHealth technology, that is, the patients might have better understood the importance of early mobilization and/or how to take care of themselves at home. This is a prime example of coproduction in which patients actively change their behavior to achieve better outcomes [31]. The patients benefiting from the mHealth technology might have been more activated because of the digital preparation and education materials when compared with patients in the nonparticipating cohort. Our theories about what patients are thinking, however, are speculative and will need to be validated by future research.

Another significant contribution of this research is demonstrating how we were able to quantify patient engagement by how often they read and responded to messages and assessing the types of messages that patients preferred [29]. There was a distinct pattern associated with the types of messages and content that patients chose to be engaged with. In this study, patients were least responsive to prompts where the hospital was asking them to do something (active); they were most engaged with emails that were fact-based informational emails (passive). Perhaps patients perceive adherence questions as overly intrusive. Future research needs to be done using mHealth technologies in a co-designed process to better discern patients’ thought processes on adherence [31].

This study has several limitations. The participating cohort consisted only of people who agreed to participate because (1) we could not ethically randomly withhold technology from patients who, based on the feedback they provided us during the co-design phase, overwhelmingly wanted to have access to and be monitored on the mHealth technology and (2) we could not ethically require nonagreeing patients to engage in substantive, recurring actions on a repeated basis—namely, reading and responding to messages every day for 50 days. Therefore, randomization was not feasible from an equipoise standpoint [32]. However, by including only patients who agreed to use mHealth technology, we acknowledge a significant methodological flaw, a selection or Berksonian bias that is often inherent in studying digital interventions—a phenomenon that arises when the sample is taken not from the general population but from a preselected subpopulation [33]. Patients who agree to participate are generally more motivated, have greater self-efficacy, are more literate, and have a variety of other attributes that make it likely that their outcomes would be better than nonagreeing patients [34-37].

To offset these limitations, we used a combination of different analytic methods, not only to address the issue of patient selection bias per se but rather to show the durability of our results and, by extension, their external generalizability to the larger population. We arrived at the same conclusions through different analytic approaches and ensured that the conclusions drawn are consistent. Specifically, we found consistent outcomes at different time intervals separated by several months. We also found that less than 1% of patients declined to participate or opted-out of participating after enrollment, and few patients indicated a preference for a language other than English in both the participating (43 patients) and nonparticipating (119 patients) surgeon groups [37].

It could also be argued that, by excluding people who preferred a language other than English, we could have unintentionally exacerbated health disparities. After all, recent research focusing on mHealth technologies has linked their implementation to enhanced risk for racial bias and health disparities, rather than adhering to the promise of equalizing health inequities [38]. However, it is important to recognize that there were too few patients in this study with a preference for a language other than English such that it would be impossible to discern differences between the populations in the nonparticipating and participating groups based on language. It would be most methodologically sound to exclude patients with a preference for a language other than English to ensure that the populations in both cohorts were similar to reduce the risk of introducing a confounding variable. Limiting this feasibility/pilot study to patients with an English preference helped control for translation variables. Furthermore, as demonstrated in Table 2, ethnicity and race were nearly
Conclusions and Future Directions

We demonstrated that patient-facing technologies that empower patients and their caregivers to become involved and informed in their care and, specifically, to play a more active role in enhancing patient care, can be effective. This is supported by the growing movement to embrace the potential of mHealth technology to transform health care outcomes, a movement of which is becoming increasingly pronounced and urgent with COVID-19 developments [7]. The purpose of this observational study was to assess whether a multicomponent mHealth technology impacts clinical outcomes and/or enhances patient-reported outcomes. We provided actionable data to demonstrate that mHealth technologies can be effective, which can help support and shape how patient-facing mHealth technology is being used during and following COVID-19.

This study should serve as a foundation for future research. Recent research has shown that technology-based implementations can be susceptible to the digital divide, which coexist with other social determinants of disparity by inadvertently masking or exacerbating racial, ethnic, or gender inequities [43]. Thus, future research needs to assess the hypothesis that mHealth technologies will improve, and certainly not widen, existing disparities by systematically examining nonrandom biases (such as health access issues) and how that might impact data.

Furthermore, there have been no studies published, to our knowledge, regarding the comparative effectiveness or efficacy of mHealth technologies [2,24]. It is currently unknown whether certain mHealth technologies, including some already in use, are effective or are more impactful than others. Further studies are also warranted to examine the impact mHealth technologies have on clinical workflows and resource utilization [24]. We demonstrated that our results were sustainable for several months, but longer research related to sustainability beyond 1 year is needed. Greater systematic evaluation and research, including prospective, multisite studies (if possible), are needed to fully characterize the effectiveness of digitally facing patient-centered technologies.

Conflicts of Interest

The MedTrak (vendor) employees who are coauthors on this manuscript helped to build the templates of the reports that allowed analyses (JS) or built the decision logic for the pathway messages that patients received (CB). CG was involved in pulling raw data and providing it to our statisticians for analyses. To ensure that the vendor did not cherry-pick data to be favorable to them, 2 independent statisticians internally conducted separate and independent spot check reviews of the databases where the study data were stored to ensure consistency in raw data. No MedTrak employees had any input in the data analyses, reviewing the data, or in drafting the manuscript.

Multimedia Appendix 1

Results from January 1, 2019 to April 30, 2019.

[DOCX File, 20 KB - mhealth_v8i6e19333_app1.docx]

References


Abbreviations

COVID-19: coronavirus disease 2019
GEE: generalized estimated equation
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
MA: medical assistant
mHealth: mobile health
OR: odds ratio
TJR: total joint replacement
Comparing a Social and Communication App, Telephone Intervention, and Usual Care for Diabetes Self-Management: 3-Arm Quasiexperimental Evaluation Study

Ching-Ju Chiu, PhD; Yung-Chen Yu, MS; Ye-Fong Du, MD, MS; Yi-Ching Yang, MD, MS; Jou-Yin Chen, MS; Li-Ping Wong, PhD; Chanuantong Tanasugarn, PhD

1Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan, Taiwan
2Department of Nursing, National Cheng Kung University Hospital, Tainan, Taiwan
3Division of Endocrinology and Metabolism, Department of Internal Medicine, National Cheng Kung University Hospital, Tainan, Taiwan
4Department of Family Medicine, National Cheng Kung University Hospital, Tainan, Taiwan
5Department of Public Health, College of Medicine, National Cheng Kung University, Tainan, Taiwan
6Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
7Faculty of Public Health, Mahidol University, Bangkok, Thailand

Corresponding Author:
Ye-Fong Du, MD, MS
Division of Endocrinology and Metabolism
Department of Internal Medicine
National Cheng Kung University Hospital
No. 138, Shengli Road, North District
Tainan, 70403
Taiwan
Phone: 886 6 2353535 ext 4577
Email: n043328@mail.hosp.ncku.edu.tw

Abstract

Background: Many technology-assisted innovations have been used to manage disease. However, most of these innovations are not broadly used by older adults due to their cost. Additionally, disease management through technology-assisted innovations has not been compared with other interventions.

Objective: In this study, we tested the employment of a free and widely used social and communication app to help older adults with diabetes manage their distress and glycemic control. We also compared the effectiveness of the app with 2 other methods, namely telephone and conventional health education, and determined which subgroup experiences the most effects within each intervention.

Methods: Adults aged ≥50 years with type 2 diabetes were recruited from Southern Taiwan (N=231) and were allocated to different 3-month interventions. Informed consent was obtained and approved by the Ministry of Science and Technology and approved by the National Cheng Kung University Hospital Institutional Review Board (No. A-ER-102-425).

Results: Participants in the mobile-based group had significant reductions in hemoglobin A1c compared with the telephone-based and usual care groups (mean changes of –0.4%, 0.1%, and 0.03%, respectively; P=.02). Diabetes-specific distress decreased to a greater extent in the mobile-based group compared to the other 2 groups (mean changes of –5.16, –3.49, and –2.44, respectively, P=.02). Subgroup analyses further revealed that the effects on reducing blood glucose levels in the social and communication app groups were especially evident in patients with lower distress scores, and diabetes-related distress was especially evident in participants who were younger than 60 years or had higher educational levels.

Conclusions: The findings of this study inform more flexible use of social and communication apps with in-person diabetes education and counselling.

(JMIR Mhealth Uhealth 2020;8(6):e14024) doi:10.2196/14024

KEYWORDS
diabetes; self-management; depression symptoms; distress; middle-aged and older adults
Introduction

Existing interventions for people with diabetes generally involve group or individual counseling for self-management education and support [1-3]. The use of technology to collect data through the internet to manage disease is increasing; many such studies focus on diabetes, including monitoring glucose or vital signs through mobile apps or websites, messaging reminders through email or SMS text messaging for self-care skills, and improving emotional status [4-8]. In recent years, mobile phones have become an increasingly important platform for the delivery of health interventions; they can be used to help individuals manage their health and disease condition [9-13]. Web-based structured integrated care has been employed by several research teams to provide information for diabetes education, record blood glucose readings, and provide personal feedback [14-16]. Many mobile apps have been developed for diabetes care. Better outcomes were observed for mobile apps that provided 2-way communication between people with diabetes and their health care team [17,18]. However, most of these interventions require the development of a structured website or a diabetes-specific mobile app; these are not translated into many languages, which limits their universal use [19].

Currently, 53% of adults age 65 and older use the Internet in Taiwan [20], and one third of older persons who are online use social networking sites such as Facebook, LinkedIn, and LINE, which were listed as age-friendly apps in 2012 [21]. Some traditional technologies (eg, telephones, video game consoles) have been utilized to conduct health-related interventions. New technologies such as exoskeleton human body posturizers [22], the Nintendo Wii console [23], and virtual reality gaming [24] have also been investigated to assist older people in improving their balance and have been shown to decrease both the fear of falling and the number of falls. For diabetes control apps in Taiwan, the National Health Insurance program developed an internet platform called My Health Bank. My Health Bank provides patients’ medical records over the past 3 years; patients can use it to record their own health data, such as BMI and blood test results, including fasting plasma glucose, hemoglobin A1c (HbA1c), and lipid profiles. Physicians can quickly understand their patients’ physical condition, treatment status, and medical examinations through this app [25]. Moreover, many studies have explored the use of new technologies such as websites or mobile apps to help with disease or health management, such as smoking cessation [26,27], medication management [28], and prevention of excessive gestational weight gain and associated maternal and child health consequences [29]. However, few existing studies have focused on the effects of existing, freely available communication apps (eg, Facebook Messenger, LINE, WhatsApp Messenger) among middle-aged and older adults; therefore, the effectiveness of social networking interventions to address the health issues of middle-aged and older adults is not certain. One of the most age-friendly and easy-to-use social network applications is known as LINE; it does not restrict the learning pace, time, or space of learners, and users also benefit from the interaction in this mobile app [30]. In addition, information about diabetes self-management can be easily transmitted through text messages, photographs, and short videos to enable people to use this information in daily life. LINE is one of the most popular apps for social communication in Taiwan [31]. However, whether this tool can be employed in self-management education and support to help older adults manage their diabetes is unknown.

According to the American Association of Diabetes Educators (AADE), 7 behavioral indicators [4] are critical to enhancing the self-care abilities of people with diabetes and to helping them reach their optimal goal of glycemic control through education and various support resources. These indicators, called the AADE7 Self-Care Behaviors, are (1) healthy eating, (2) being active, (3) monitoring, (4) taking medication, (5) problem solving, (6) reducing risks, and (7) healthy coping. The results of one study indicated that most diabetes apps do not adhere to more than 2 AADE7 self-care behavior guidelines [32]. Therefore, the aims of this study were to test 3 intervention modes that were developed based on the AADE7 self-care behavior guidelines, including mobile-based, telephone-based, and usual care interventions. We compared the impact of the mobile-based, telephone-based, and usual care interventions on diabetes-related distress, depressive symptoms, and glycemic control in middle-aged and older adults living in the community. We also compared the degrees of acceptance within the 3 groups to determine which subgroup experienced the most effects.

Methods

This study was a 3-arm quasieperimental design. We recruited participants aged ≥50 years with type 2 diabetes through a medical chart review. Exclusion criteria were people with renal dysfunction (creatinine ≥1.5 milligrams per deciliter) or who were receiving dialysis treatment. People who were diagnosed with dementia, cognitive impairment, or major depression were also excluded. We recruited our participants from the endocrine and family medicine clinics of a medical center in Southern Taiwan. Informed consent was obtained at the Ministry of Science and Technology, and the study was approved by the National Cheng Kung University Hospital Institutional Review Board (No. A-ER-102-425).

In the mobile-based group, participants were required to own a mobile device and to be capable of using the mobile device (n=49). Participants who did not have a mobile device and access to the internet at home were randomized to either the telephone-based interview group (n=91) or the usual care group (n=91).

Each group was provided with diabetes-related self-care information and provided with emotional support through usual care. In the mobile-based intervention group, in addition to the participants’ usual care, we used a communication app called LINE to send multimedia messages about diabetes self-management for 12 weeks. LINE is a free instant communication service. In Taiwan, over 90% of people install LINE on their mobile devices [33]. LINE users can set up their own accounts using their telephone number or email address and can link with friends by telephone number or LINE ID. When the participants were ready to join our study, we checked their telephone number or LINE ID and linked it to our research.
account, and a welcome message was sent to the individual accounts of the participants. The research team verified whether the messages we sent to each individual account were marked as “read.” The participants were also able to interact with our research team by LINE. Messages were sent 3 times a week, and the participants received an average of 5 messages per week. Structural diabetes health education modules were developed according to AADE7 guidelines. We included 6 of the behavioral indicators in our messages: healthy eating (eg, how to select appropriate food and observe amounts of sweeteners); being active (eg, suitable ways to exercise and precautions related to exercise); monitoring (eg, the standard glucose values before and after a meal); taking medication (eg, what to do if they forgot to take their medicine); (5) problem solving (eg, how to relax and search for health information); and (6) reducing risk (eg, foot care skills). The topics of the weekly messages are shown in Textbox 1 and sample messages are shown in Multimedia Appendix 1.

In the telephone-based intervention group, the participants received 3-4 phone calls lasting 30-60 minutes each from a diabetes health educator. The first call explored the participant’s diet, sleep habits, exercise level, and blood glucose control as well as their general and diabetes-specific health conditions. The people with diabetes were asked to talk about their feelings and lifestyle changes after they were diagnosed with diabetes. The goal was to understand their daily routines, thoughts, worries, related feelings, and behavior. According to the baseline assessment, in the second and subsequent phone calls, the assistants asked the participants to discuss self-care problems related to diabetes and their feelings about the disease in addition to the AADE7 modules. The usual care group received a routine intervention for people with diabetes in hospital, where the participants received 5-10 minutes of usual care (including a nutrition consultation, exercise guidance, medication, and other self-care skills related to diabetes) once every 3 months by family medicine physicians in family medicine outpatient departments and by certified diabetes educators at diabetic health education clinics.

We collected data on the health conditions of the participants, such as the year they were diagnosed with diabetes, body weight, height, anti-diabetic therapy, and recent HbA1c levels through a medical chart review. We reviewed the medical record data related to complications such as diabetic foot, hypertension, and stroke. We also collected self-reported data about sociodemographic characteristics, health behavior, and psychological well-being. Sociodemographic factors included age, sex, education, ethnicity, and marital status. The educational levels were divided into 6th grade and below, grades 7-12, and grades 13 and above. The options in the ethnicity category were Taiwanese, mainlander, Hakka, and aborigine. The possible answers to the question on marital status were widowed, married and living with spouse, and other (divorced, separated, or unmarried). The health behavior variables included the frequency and time of exercise, dietary habits, tobacco use, alcohol use, and the amount of alcohol consumed. These data were collected at baseline and after the intervention. Diabetes-specific emotional distress was measured using the Problem Areas in Diabetes (PAID) scale. It contains 20 questions; each question describes a problem that a person with diabetes may have. The PAID scale uses a 5-point item scale: 1=Not a problem, 2=Minor problem, 3=Moderate problem, 4=Somewhat serious problem, and 5=Serious problem. A higher PAID score indicates greater emotional distress related to diabetes. The total scale of PAID is 0-100, and the commonly used cutoff score of ≥40 was used to indicate elevated levels of diabetes-specific emotional distress [34].

Depressive symptoms were measured with the Center for Epidemiological Studies Depression Scale (CES-D), which is a self-reported questionnaire. The CES-D contains 20 questions. It has been found to have good validity and reliability for different races and ages [35]. In the pretest, we used the brief CES-D. The brief CES-D contains 10 questions on a 4-point scale to classify the depression status of older persons in the most recent week, with possible answers of “not at all,” “usually,” and “always.” The short-form CES-D scale ranges from 0-30, with a higher CES-D score indicating a higher level of depression. The cutoff scores for depressive symptoms were ≥16 for the full-length questionnaire and ≥10 for the 10-item version for older adults [36,37].
Textbox 1. The topics of the weekly SMS text messages.

**Week 1. Healthy eating**
- Regular diet and principles of eating out for people with diabetes
- Appropriate food and desserts for people with diabetes
- Awareness of artificial sweeteners
- Low calorie diet education

**Week 2. Healthy eating**
- Recording daily amounts of milk, starch, vegetables, fruits, beans, and meat
- Carbohydrate counting
- Glycemic indices of common fruits and low fat food groups

**Week 3. Being active**
- The importance and principles of exercise for people with diabetes
- Photographs of a variety of exercise positions for middle-aged and older adults, such as warmups and calisthenics

**Week 4. Being active**
- Photographs of a variety of exercise positions, such as those promoting strength, endurance, and flexibility

**Week 5. Taking medication**
- Insulin group:
  - Myths and facts about insulin injections
  - Introduction to different types of insulin and their durations of use
- Oral drugs group:
  - Directions and reminders for taking medication
  - Tips on what to do if they forget to take their medication

**Week 6. Taking medication**
- Insulin group:
  - Tips on injection skills and equipment
  - Travel guidelines for people with diabetes
- Oral drugs group:
  - Myths and facts about taking oral drugs
  - Travel guidelines for people with diabetes

**Week 7. Monitoring**
- The importance of glycemic monitoring
- Appropriate times and frequency of monitoring
- The meaning of fasting glucose

**Week 8. Monitoring**
- Hypoglycemia and hyperglycemia symptoms
- Common causes of hypoglycemia and hyperglycemia
- Tips for preventing hypoglycemia and hyperglycemia

**Week 9. Reducing risk**
- Causes of diabetic foot
Results

The baseline characteristics of the 3 groups in our study are shown in Table 1. No significant differences were observed in terms of gender, marital status, antidiabetic therapy, eye problems, foot problems, hypertension, lung disease, heart disease, cancer, arthritis, smoking, or exercise behavior across the 3 groups; also, no significant differences were observed in the means of the duration of diabetes, HbA1c level, BMI, or CES-D score. However, participants in the mobile-based intervention group (n=49) were significantly younger and had higher educational levels than those in the telephone-based group (n=91) and usual care group (n=91). Also, participants in the mobile-based intervention group had higher PAID scores than those in the other 2 groups at baseline. These factors were used as stratification variables in our subgroup analysis.

Table 2 presents the pretest and posttest HbA1c, CES-D, and PAID scores and the differences between them across the 3 groups. Participants in the mobile-based group showed significant improvement in their HbA1c levels after the intervention (posttest to pretest: –0.40%, P=.04) and PAID scores (posttest to pretest: –5.16, P=.01). We further used the Kruskal-Wallis test to examine the differences in the main outcomes across the groups. Our results showed that changes in the participants’ HbA1c levels in the mobile-based intervention group (posttest to pretest: –0.40%) were significantly higher than those in the telephone-based group (posttest to pretest: 0.10%) and the usual care group (posttest to pretest: 0.03%). Also, the absolute improvement of the PAID score in the mobile-based intervention group (posttest to pretest: –5.16) was significantly higher than that of the telephone-based group (posttest to pretest: –3.49) and the usual care group (posttest to pretest: –2.44).

Due to the significant differences shown in Table 1 and the possible confounding associated with age, educational level, and depression symptoms at baseline, we further conducted a sensitivity analysis that stratified the participants by age, educational level, and baseline PAID score (Table 3). It was found that for participants aged <60 years, diabetes distress was significantly reduced in both the mobile-based intervention group (score –7.45, P=.01) and the telephone-based intervention group (score –4.87, P=.008). For participants with educational levels ≥7 grades, diabetes distress was also significantly reduced in both the mobile-based intervention group (score –8.80, P=.001) and the telephone-based intervention group (score –7.44, P=.008). In addition, participants in the mobile-based intervention group experienced a greater reduction in diabetes distress than those in the telephone-based intervention group, regardless of subgroup: age <60 years (the differences in the PAID score were mobile –7.45; telephone –4.87; usual care –4.14) or education ≥7th grade (the differences in the PAID score were mobile –8.80; telephone –7.44; usual care –5.00). Furthermore, we found that if the participants’ PAID was above the average (3.75), diabetes distress was significantly reduced in the mobile-based (score –6.89, P<.001), telephone-based (score –9.01, P<.001), and usual care groups (score –7.90, P=.001) after the intervention. In addition, for participants with diabetes distress scores lower than the median, only the mobile-based intervention group exhibited significantly reduced HbA1c levels (score –0.31%, P=.04).
Table 1. The baseline sociodemographic, clinical, and psycho-behavioral characteristics of the mobile-based, telephone-based, and usual care groups (N=231).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mobile-based group (n=49)</th>
<th>Telephone-based group (n=91)</th>
<th>Usual care group (n=91)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>58.6 (6.0)</td>
<td>64.70 (8.30)</td>
<td>64.70 (9.50)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male gender, n (%)</td>
<td>32 (65)</td>
<td>50 (55)</td>
<td>44 (48)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Education level, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 6 and below</td>
<td>21 (23)</td>
<td>32 (35)</td>
<td>22 (45)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Grades 7-12</td>
<td>71 (82)</td>
<td>71 (78)</td>
<td>75 (82)</td>
<td>.16</td>
</tr>
<tr>
<td>Married or partnered, n (%)</td>
<td>43 (88)</td>
<td>71 (78)</td>
<td>75 (82)</td>
<td>.16</td>
</tr>
<tr>
<td><strong>Diabetes and other clinical variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of diabetes (years), mean (SD)</td>
<td>7.5 (6.1)</td>
<td>10.60 (8.60)</td>
<td>10.40 (8.18)</td>
<td>.10</td>
</tr>
<tr>
<td><strong>Diabetes treatment, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet and exercise</td>
<td>7 (14)</td>
<td>5 (6)</td>
<td>10 (11)</td>
<td>.11</td>
</tr>
<tr>
<td>Oral medication</td>
<td>31 (63)</td>
<td>58 (63)</td>
<td>71 (78)</td>
<td>.71</td>
</tr>
<tr>
<td>Insulin</td>
<td>11 (23)</td>
<td>20 (23)</td>
<td>10 (11)</td>
<td>.71</td>
</tr>
<tr>
<td>HbA1c level (%)</td>
<td>7.6 (1.5)</td>
<td>7.6 (1.6)</td>
<td>7.6 (1.3)</td>
<td>.94</td>
</tr>
<tr>
<td>BMI (kilograms per square meter), mean (SD)</td>
<td>25.9 (3.7)</td>
<td>25.80 (3.30)</td>
<td>25.90 (4.40)</td>
<td>.10</td>
</tr>
<tr>
<td>Eye problems, n (%)</td>
<td>4 (8)</td>
<td>7 (8)</td>
<td>10 (11)</td>
<td>.10</td>
</tr>
<tr>
<td>Foot problems, n (%)</td>
<td>2 (4)</td>
<td>10 (11)</td>
<td>5 (6)</td>
<td>.22</td>
</tr>
<tr>
<td>Hypertension, n (%)</td>
<td>19 (39)</td>
<td>49 (54)</td>
<td>45 (50)</td>
<td>.23</td>
</tr>
<tr>
<td>Lung disease, n (%)</td>
<td>0 (0)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>.15</td>
</tr>
<tr>
<td>Heart disease, n (%)</td>
<td>9 (18)</td>
<td>11 (12)</td>
<td>7 (8)</td>
<td>.17</td>
</tr>
<tr>
<td>Cancer, n (%)</td>
<td>2 (4)</td>
<td>6 (7)</td>
<td>8 (9)</td>
<td>.57</td>
</tr>
<tr>
<td>Arthritis, n (%)</td>
<td>2 (4)</td>
<td>10 (11)</td>
<td>8 (9)</td>
<td>.38</td>
</tr>
<tr>
<td><strong>Psychobehavioral variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking, n (%)</td>
<td>4 (8)</td>
<td>13 (14)</td>
<td>4 (4)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Exercise, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>4 (9)</td>
<td>18 (20)</td>
<td>15 (17)</td>
<td>.09</td>
</tr>
<tr>
<td>1-2 times or less than 90 minutes per week</td>
<td>16 (32)</td>
<td>17 (19)</td>
<td>13 (14)</td>
<td></td>
</tr>
<tr>
<td>More than 3 times and &gt;90 minutes per week</td>
<td>29 (60)</td>
<td>56 (61)</td>
<td>63 (69)</td>
<td></td>
</tr>
<tr>
<td>CES-Db score (0-30), mean (SD)</td>
<td>6.8 (5.1)</td>
<td>5.20 (3.90)</td>
<td>5.60 (4.80)</td>
<td>.14</td>
</tr>
<tr>
<td>Diabetes-specific emotional distress score (PAIDc, 0-100), mean (SD)</td>
<td>13.5 (12.4)</td>
<td>5.40 (7.40)</td>
<td>5.20 (9.30)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*aHbA1c: hemoglobin A1c*  
*bCES-D: Center for Epidemiological Studies Depression Scale*  
*cPAID: Problem Areas in Diabetes Scale*
Table 2. Pretest and posttest HbA\textsubscript{1c}, CES-D, and PAID scores across the 3 groups and the differences between them.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mobile-based group</th>
<th>Telephone-based group</th>
<th>Usual care group</th>
<th>Kruskal-Wallis test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Chi-square (2)</td>
</tr>
<tr>
<td></td>
<td>P value\textsuperscript{a}</td>
<td>P value\textsuperscript{a}</td>
<td>P value\textsuperscript{a}</td>
<td>P value</td>
</tr>
<tr>
<td>HbA\textsubscript{1c}\textsuperscript{b} (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>7.8 (1.49)</td>
<td>7.64 (1.56)</td>
<td>7.67 (1.28)</td>
<td>8.3</td>
</tr>
<tr>
<td>Posttest</td>
<td>7.18 (1.05)</td>
<td>7.74 (1.47)</td>
<td>7.70 (1.18)</td>
<td>.02</td>
</tr>
<tr>
<td>Difference\textsuperscript{c}</td>
<td>-0.40 (1.24)</td>
<td>0.10 (1.12)</td>
<td>0.03 (1.1)</td>
<td></td>
</tr>
<tr>
<td>CES-D\textsuperscript{d} score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>6.78 (5.08)</td>
<td>5.19 (3.91)</td>
<td>5.59 (4.75)</td>
<td>4.0</td>
</tr>
<tr>
<td>Posttest</td>
<td>7.95 (5.76)</td>
<td>3.84 (2.40)</td>
<td>4.47 (3.09)</td>
<td>.13</td>
</tr>
<tr>
<td>Difference\textsuperscript{c}</td>
<td>0.98 (6.14)</td>
<td>-1.34 (4.16)</td>
<td>-1.12 (4.86)</td>
<td></td>
</tr>
<tr>
<td>PAID\textsuperscript{e} score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>13.52 (12.40)</td>
<td>5.44 (7.39)</td>
<td>5.24 (5.29)</td>
<td>7.6</td>
</tr>
<tr>
<td>Posttest</td>
<td>9.31 (9.79)</td>
<td>2.06 (6.14)</td>
<td>2.90 (4.79)</td>
<td>.02</td>
</tr>
<tr>
<td>Difference\textsuperscript{c}</td>
<td>-5.16 (12.58)</td>
<td>-3.49 (9.97)</td>
<td>-2.44 (10.60)</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Calculated using the Kruskal-Wallis test.

\textsuperscript{b}HbA\textsubscript{1c}: hemoglobin A1c.

\textsuperscript{c}Difference: posttest value – pretest value.

\textsuperscript{d}CES-D: Center for Epidemiological Studies Depression Scale.

\textsuperscript{e}PAID: Problem Areas in Diabetes Scale.
### Table 3. Sensitivity analysis comparing the intervention effects in the different subgroups.

<table>
<thead>
<tr>
<th>Subgroup and effect</th>
<th>Mobile-based group</th>
<th>Telephone-based group</th>
<th>Usual care group</th>
<th>Kruskal-Wallis test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Score</td>
<td>( P ) value</td>
<td>n</td>
</tr>
<tr>
<td>Age ≥60 years (n=138)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>19</td>
<td>-0.42</td>
<td>.09</td>
<td>61</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>2.35</td>
<td>.21</td>
<td></td>
<td>-1.44</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>-2.06</td>
<td>.49</td>
<td></td>
<td>-2.82</td>
</tr>
<tr>
<td>Age &lt;60 years (n=93)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>30</td>
<td>-0.38</td>
<td>.19</td>
<td>30</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>-0.04</td>
<td>.97</td>
<td></td>
<td>-1.14</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>-7.45</td>
<td>.01</td>
<td></td>
<td>-4.87</td>
</tr>
<tr>
<td>Education≥7 grades (n=69)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>26</td>
<td>-0.25</td>
<td>.31</td>
<td>22</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>0.09</td>
<td>.94</td>
<td></td>
<td>-0.71</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>-8.80</td>
<td>.001</td>
<td></td>
<td>-7.44</td>
</tr>
<tr>
<td>Education&lt;7 grades (n=162)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>23</td>
<td>-0.57</td>
<td>.07</td>
<td>69</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>2.18</td>
<td>.20</td>
<td></td>
<td>-1.54</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>-0.22</td>
<td>.95</td>
<td></td>
<td>-2.26</td>
</tr>
<tr>
<td>PAID mean score ≥3.75 (n=34)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>17</td>
<td>-0.41</td>
<td>.08</td>
<td>6</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>1.27</td>
<td>.28</td>
<td></td>
<td>-1.23</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>-6.89</td>
<td>.005</td>
<td></td>
<td>-9.01</td>
</tr>
<tr>
<td>PAID mean score &lt;3.75 (n=197)</td>
<td></td>
<td></td>
<td></td>
<td>n</td>
</tr>
<tr>
<td>( \Delta \text{HbA}_1c ) (%)</td>
<td>32</td>
<td>-0.31</td>
<td>.04</td>
<td>85</td>
</tr>
<tr>
<td>( \Delta \text{CES-D} )</td>
<td>-0.43</td>
<td>.74</td>
<td></td>
<td>-1.44</td>
</tr>
<tr>
<td>( \Delta \text{PAID} )</td>
<td>3.04</td>
<td>.10</td>
<td></td>
<td>1.78</td>
</tr>
</tbody>
</table>

\( a \): Hemoglobin A\(_1c\)

\( b \): Center for Epidemiological Studies Depression Scale.

\( c \): Problem Areas in Diabetes Scale.

---

**Discussion**

**Principal Findings**

This study examined the effectiveness of a mobile-based intervention on the glycemic control of people with diabetes, depressive symptoms, and diabetes-specific distress in comparison with telephone-based intervention and usual care. The results provide evidence that the mobile-based intervention has the greatest potential to improve glycemic control and reduce diabetes-specific distress among the 3 interventional methods. For participants in the mobile-based group who were less than 60 years of age or who had higher educational levels at baseline, their distress scores decreased more dramatically than those of the participants in the other 2 groups after the intervention. In addition, for participants with distress scores below the average at baseline, those who received the intervention based on the social and communication app LINE had significantly improved HbA\(_1c\) values, which was not observed in the telephone-based group or usual care group.

Previous studies evaluating technology-enabled self-management education mostly used web-based structured diabetes education, time-consuming nurse coaching, and online educational discussion groups to aid diabetes self-management [8]. Although these approaches can reduce depressive symptoms and diabetes-related distress, they are labor-intensive and may be expensive. In the present study, we utilized a patient-centered approach and a freely available social app. We sent text and photographic messages based on the AADE’s 7 indicators for diabetes self-management to middle-aged and older adults through the free social and communication app LINE, which...
the participants were familiar with. In addition, we provided social support, problem solving, and communication with professionals through the LINE app. With figure-rich messages for diabetic education, the LINE app is suggested to be an effective platform to execute diabetes education and support; it may improve diabetes glycemic control and reduce diabetes-related distress more than usual care, supporting previous studies indicating the strength of mobile phones in this area [38]. The strengths of social and communication apps include their ability to send multimedia messages and free video calls; also, using mobile-based devices, participants can browse their messages repeatedly and magnify the image-based messages, which addresses the decline in cognitive functioning required for older adults to manage their disease. Also, the bright colors and clear font size in the photo messages are age-friendly and help older adults avoid eye fatigue [39]. We found that the participants not only experienced decreased diabetes distress but also had increased glycemic control performance.

The success of the application of the LINE app in diabetes care in middle-aged and older adults in Taiwan may be due to their cultural background of the participants and the readily accessible and affordable health care system. As addressed in previous literature, preferences for mobile apps developed for diabetes management are different in different countries; Americans prefer apps that support their decision-making, while Chinese and Middle Eastern residents prefer app-based communication functions [40,41]. Korean people also benefited from an automated self-measured blood glucose upload built-in diabetes self-management app with physician response once a week [42].

In brief, patients in Asia and the Middle East prefer to seek advice from their health care teams directly through the communication app, while patients in western countries prefer to solve problems by themselves if the diabetes-specific app can provide adequate resources to support their decision-making. On the other hand, the language of most diabetes apps is English (85.4%); this limits the selection of diabetes apps in Taiwan. Additionally, patients prefer free apps to paid apps [43]. In this era of knowledge explosion, patients can search for diabetes management articles online; however, they may not recognize whether the information in these articles is correct [44]. There is no one app that is suitable for all patients with diabetes.

Development of diabetic self-management mobile apps in Taiwan also requires much user feedback [45]. However, even if a dedicated developed app that contains all suggested functions and articles beneficial for diabetes self-management were accessible, patients would not use all the built-in functions to enhance their self-care behavioral skills [46]. For the reasons addressed above, the application of LINE for diabetes care in Taiwan is appropriate in many aspects; it can be used to communicate with health care teams, can provide carefully selected articles that assist diabetes self-management, and is freely available. We also used a strategy of not providing too many messages at a time; therefore, participants could absorb the diabetes self-management knowledge in steps during the 12-week intervention period.

In the present study, we found when the diabetes-related distress scores for the participants were higher before the intervention, their diabetes distress was significantly reduced in both the mobile-based and telephone-based groups. Increasing numbers of studies are focusing on mobile-related devices or applications rather than telephone-based interventions due to the fact that such devices and applications are cost-effective, can reach large populations, suggest promising outcomes, and have expanded rapidly in the past decade [47-50]. This study also suggests that mobile-related applications are more cost-effective than telephone-based interventions and are very efficient. Previous studies have provided services that deliver interventions, including brief psychological therapies, mental health assessments, psychotropic medications, and social support, enhanced by patient-led case conferences aiming to optimize diabetes care. The results of these studies indicate a significant reduction in HbA1c of 3.5%, which was associated with reductions at 1-year follow-up [51]. Our study found that for the mobile-based intervention group, when we focused on people with diabetes whose diabetes distress scores were lower than the median score, there was a significant reduction in HbA1c scores during the 12-week period. This can be explained by the fact that information seeking is one of the most important elements of coping strategies of chronic illness among patients who have comorbid depression [52]; immediate responses and conversation privacy are additional strengths of mobile-based interventions.

In this study, we found that the decrease of the PAID score with the mobile-based intervention was greater when the participants were younger and when their education levels were higher. This finding echoes those of previous studies indicating that people newly diagnosed with diabetes are willing to participate in self-management programs where their medical outcomes are being effectively targeted [53,54].

**Limitations**

There were some limitations of this study. First, this study had a small sample size. Second, it did not randomly allocate participants to the 3 interventions. However, we compared the mobile-based group with the telephone-based and usual care groups and found manageable differences. Third, this study did not evaluate changes in behavior-related variables. Future studies are suggested to test the effects of this intervention in a broader sample and to evaluate both its behavioral and long-term effects.

**Conclusion**

Existing free social and communication apps are effective to improve glycemic control and reduce diabetes-specific distress in older adults with diabetes. For participants younger than 60 years, with higher educational levels, or with higher diabetes-related distress at baseline, the social and communication app reduced their distress scores more dramatically than those of participants in the other 2 groups after the intervention. In addition, for participants whose distress scores were below average at baseline, the social and communication app significantly improved their glycemic control.
Acknowledgments
This work was supported by grant MOST-103-2314-B-006-038-MY3 from the Taiwan National Science Council. The authors appreciate Dr. Ye-Fong Du for helping recruit the patients with diabetes for this study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Sample messages.

References


42. Yang Y, Lee EY, Kim H, Lee S, Yoon K, Cho J. Effect of a Mobile Phone-Based Glucose-Monitoring and Feedback System for Type 2 Diabetes Management in Multiple Primary Care Clinic Settings: Cluster Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16266 [FREE Full text] [doi: 10.2196/16266] [Medline: 32130172]

43. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]


Abbreviations

AADE: American Association of Diabetes Educators
HbA1c: hemoglobin A1c
PAID: Problem Areas in Diabetes Scale
CES-D: Center for Epidemiological Studies Depression Scale
Digital Phenotyping Self-Monitoring Behaviors for Individuals With Type 2 Diabetes Mellitus: Observational Study Using Latent Class Growth Analysis

Qing Yang1, PhD; Daniel Hatch1, PhD; Matthew J Crowley2,3, MD; Allison A Lewinski2, PhD; Jacqueline Vaughn1, BSN; Dori Steinberg1, PhD; Allison Vorderstrasse4, PhD; Melin Jiang5, MSc; Ryan J Shaw1,6, PhD

1School of Nursing, Duke University, Durham, NC, United States
2Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Medical Center, Duke University, Durham, NC, United States
3Division of Endocrinology, Diabetes and Metabolism, School of Medicine, Duke University, Durham, NC, United States
4College of Nursing, New York University, New York, NY, United States
5Department of Biostatistics, University of Florida, Gainesville, FL, United States
6Center for Applied Genomics and Precision Medicine, School of Medicine, Duke University, Durham, NC, United States

Corresponding Author:
Qing Yang, PhD
School of Nursing
Duke University
307 Trent Drive
Durham, NC, 27710
United States
Phone: 1 9196139768
Email: qing.yang@duke.edu

Abstract

Background: Sustained self-monitoring and self-management behaviors are crucial to maintain optimal health for individuals with type 2 diabetes mellitus (T2DM). As smartphones and mobile health (mHealth) devices become widely available, self-monitoring using mHealth devices is an appealing strategy in support of successful self-management of T2DM. However, research indicates that engagement with mHealth devices decreases over time. Thus, it is important to understand engagement trajectories to provide varying levels of support that can improve self-monitoring and self-management behaviors.

Objective: The aims of this study were to develop (1) digital phenotypes of the self-monitoring behaviors of patients with T2DM based on their engagement trajectory with multiple mHealth devices, and (2) assess the association of individual digital phenotypes of self-monitoring behaviors with baseline demographic and clinical characteristics.

Methods: This longitudinal observational feasibility study included 60 participants with T2DM who were instructed to monitor their weight, blood glucose, and physical activity using a wireless weight scale, phone-tethered glucometer, and accelerometer, respectively, over 6 months. We used latent class growth analysis (LCGA) with multitrajectory modeling to associate the digital phenotypes of participants’ self-monitoring behaviors based on their engagement trajectories with multiple mHealth devices. Associations between individual characteristics and digital phenotypes on participants’ self-monitoring behavior were assessed by analysis of variance or the Chi square test.

Results: The engagement with accelerometers to monitor daily physical activities was consistently high for all participants over time. Three distinct digital phenotypes were identified based on participants’ engagement with the wireless weight scale and glucometer: (1) low and waning engagement group (24/60, 40%), (2) medium engagement group (20/60, 33%), and (3) consistently high engagement group (16/60, 27%). Participants that were younger, female, nonwhite, had a low income, and with a higher baseline hemoglobin A1c level were more likely to be in the low and waning engagement group.

Conclusions: We demonstrated how to digitally phenotype individuals’ self-monitoring behavior based on their engagement trajectory with multiple mHealth devices. Distinct self-monitoring behavior groups were identified. Individual demographic and clinical characteristics were associated with different self-monitoring behavior groups. Future research should identify methods to provide tailored support for people with T2DM to help them better monitor and manage their condition.
Introduction

Sustained self-management with consistent self-monitoring is essential for individuals with type 2 diabetes mellitus (T2DM) to help them maintain optimal health [1]. Mobile health (mHealth) devices (eg, apps, Fitbit, Apple Watch, wireless scale, glucometer) are widely available and can help support engagement in T2DM self-management behaviors [2]. Using mHealth devices to monitor weight, blood glucose levels, activity levels, and dietary behaviors has proven to be feasible and effective in adults with T2DM [3-5]. Despite these benefits of mHealth tools, research indicates that engagement with mHealth tools decreases over time, and these trends also vary according to individual characteristics [6-10]. Determining these patterns of engagement with mHealth tools over time and how individual characteristics are associated with various patterns may provide crucial understanding on the use of mHealth tools to support T2DM self-monitoring and self-management.

Digital phenotyping, the concept of using data from mHealth devices to augment assessment of human illness, is rapidly emerging [11]. To date, digital phenotyping has been successfully used to track behavior and symptom data and to refine diagnosis and risk prediction for psychiatric disorders [12], dementia [13], and asthma [14]. Digital phenotyping has also been used to facilitate chronic disease management, such as using wearable accelerometers to track functional outcomes in patients with neurological disorders and to facilitate rehabilitation programs [15]. Although the actual readings or values from mHealth devices provide vital information for disease diagnosis, prognosis, and management, the engagement trajectories with multiple mHealth devices over time also provide crucial information about self-monitoring behaviors for patients with chronic diseases. Individuals who engage with mHealth devices more frequently indicate better self-monitoring behavior.

Latent class growth analysis (LCGA) is a type of growth mixture model that can determine individual phenotypes by identifying subgroups who follow similar trajectories over time on one or more outcomes. LCGA has been used extensively in the social sciences [16]. In medical and nursing research, LCGA has been used to quantify patient risk profiles based on physiological measures [17] and symptom research [18]. The method was extended by Jones and Nagin [19] to identify distinct subgroups based on trajectories across multiple outcomes.

In this study, we sought to demonstrate how to digitally phenotype the self-monitoring behaviors of individuals with T2DM based on their engagement trajectories with multiple mHealth devices using LCGA with multitrajectory modeling. Further, we explored if the participants’ digital phenotypes on self-monitoring behaviors varied according to their baseline demographic and clinical characteristics.

Methods

Design and Sample

This was a longitudinal observational feasibility study using multiple mHealth devices for patients with T2DM. The complete details of the study protocol were previously reported [5]. Sixty individuals with T2DM were recruited from a single primary care clinic and were followed for 6 months. As described in Table 1, participants were asked to perform self-monitoring using three measures on three mobile devices provided by the study over 6 months. This included weight (pounds) measured by a cellular-enabled scale provided by BodyTrace (Palo Alto, CA, USA), blood glucose (mg/dL) measured through a phone-tethered glucometer provided by iHealth (Mountain View, CA, USA), and physical activity measured in daily steps by a wrist-worn accelerometer and associated fitness app provided by Fitbit (San Francisco, CA, USA). Participants reported demographic information at baseline in Research Electronic Data Capture, a secure, web-based software platform designed to support data capture for research studies [20,21]. Daily monitoring on weight and physical activity were required by the study protocol, whereas blood glucose monitoring was prescribed by the primary care physician, which was performed at least once a week. Participants’ hemoglobin A1c (HbA1c) values were extracted from their electronic health record from the closest date to baseline and 6 months postbaseline. Duke University’s Institutional Review Board approved all study activities (IRB No. Pro00071569).

Table 1. Mobile health devices used and time points for data collection.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Instrument</th>
<th>Time points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight (pounds)</td>
<td>Cellular-enabled Scale (BodyTrace)</td>
<td>Daily</td>
</tr>
<tr>
<td>Blood glucose (mg/dL)</td>
<td>Food and Drug Association-approved wireless glucometer</td>
<td>As prescribed by the primary care physician, at least once a week</td>
</tr>
<tr>
<td>Physical activity (number of steps)</td>
<td>Triaxial accelerometer and associated fitness app (Fitbit)</td>
<td>Daily</td>
</tr>
<tr>
<td>Hemoglobin A1c (mmol/mol)</td>
<td>Electronic health record laboratory results</td>
<td>Baseline and 6 months postbaseline</td>
</tr>
</tbody>
</table>
Measures

Self-monitoring behaviors were captured by engagement with different mHealth devices on tracking weight, blood glucose, and physical activity. We operationalized engagement with the wireless weight scale and accelerometer as the percentage of days that the participants used the devices during 13 biweekly periods over 6 months. Since some participants may measure blood glucose multiple times per day, we operationalized engagement with the glucometer as the percentage of days that participants measured blood glucose at least once a day.

Covariates included age (years), gender, race (nonwhite, white), income (1=<US $10,000, 2=US $10,000-19,999, 3=US $20,000-29,999, 4=US $30,000-39,999, 5=US $40,000-49,999, 6=US $50,000-59,999, 7=US $60,000-79,999, 8=US $80,000), insulin dependence (currently taking any insulin medication), and baseline HbA1c values.

Data Analysis

Descriptive statistics were used to summarize demographic and other participant characteristics at baseline, including age, gender, race, income, insulin dependence, and HbA1c level. Empirical summary plots of biweekly engagement rates over 6 months were created for each device to illustrate the trajectories of self-monitoring behaviors for weight, blood glucose, and physical activity.

We conducted LCGA using SAS Proc Traj [19,22] to identify latent groups of trajectories in biweekly engagement over the 6-month observation period. We first modeled the trajectories of biweekly engagement of each device separately to determine the number of latent classes that offered the best fit for each device. Because engagement rate is a continuous variable with an approximately normal distribution, we used the censored normal distribution (cnorm). Based on the empirical summary plots, we tested both linear and quadratic trend models and chose the number of latent groups based on different number of groups using both the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values in addition to clinical judgement of the study team. After modeling the trajectories for each device, we modeled the trajectories of adherence to the devices simultaneously using a multitrajectory model.

To examine the relationships between participant characteristics, clinical variables, and latent trajectory group membership, we conducted analysis of variance (ANOVA) for age, income level, and baseline HbA1c value, and Chi square tests for race, gender, and insulin dependence. Finally, ANOVA was conducted to assess if a latent trajectory group identified in the LCGA multitrajectory model was associated with 6-month HbA1c values and changes in HbA1c from baseline to 6 months. All data analyses were conducted with SAS 9.4 (Cary, NC, USA).

Results

Sample

Demographic characteristics are presented in Table 2. The majority of the participants were women and nonwhite. The median income was US $40,000-49,999, and approximately half of the participants were currently using insulin medication. More detailed information on the sample and recruitment was reported previously [10].

Table 2. Demographic characteristics (N=60).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>55.1 (11.7)</td>
</tr>
<tr>
<td>Gender: Female, n (%)</td>
<td>43 (72)</td>
</tr>
<tr>
<td>Race: Black/Non-White, n (%)</td>
<td>39 (65)</td>
</tr>
<tr>
<td>Income level (USD), n (%)</td>
<td></td>
</tr>
<tr>
<td>&lt;$10,000</td>
<td>4 (7)</td>
</tr>
<tr>
<td>$10,000-19,999</td>
<td>3 (5)</td>
</tr>
<tr>
<td>$20,000-29,999</td>
<td>8 (14)</td>
</tr>
<tr>
<td>$30,000-39,999</td>
<td>5 (9)</td>
</tr>
<tr>
<td>$40,000-49,999</td>
<td>11 (20)</td>
</tr>
<tr>
<td>$50,000-59,999</td>
<td>6 (11)</td>
</tr>
<tr>
<td>$60,000-79,999</td>
<td>5 (9)</td>
</tr>
<tr>
<td>$80,000</td>
<td>14 (25)</td>
</tr>
<tr>
<td>Insulin dependent, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>29 (48)</td>
</tr>
<tr>
<td>No</td>
<td>31 (52)</td>
</tr>
<tr>
<td>Hemoglobin A1c at baseline, mean (SD)</td>
<td>8.1 (1.8)</td>
</tr>
</tbody>
</table>
Engagement Trajectories and Self-Monitoring Behavior Phenotypes

Empirical summary plots of engagement with different mHealth devices are presented in Figure 1 to show overall engagement trends. The average engagements with the wireless scale and glucometer for all participants were moderate (52%-72%) and showed a slightly decreasing trend for the first 2 weeks. In contrast, engagements with the accelerometers remained high over time at around 90% and with very minimal variability across all participants.

Three latent classes of engagement trajectories were identified (Table 3, Figure 2): low and waning engagement group, medium engagement group, and consistently high engagement group. The AIC and BIC values of this model were –825.1 and –846.0, respectively. In the low engagement group, individuals had relatively lower engagement with daily weight and glucose monitoring at baseline (40% and 56%, respectively) and showed a statistically significant decrease in daily weight and glucose monitoring over time. The drop in engagement with daily weight monitoring was faster in the first 2 weeks and was captured by a significant quadratic term. The moderate engagement group showed moderate engagement with daily weight and glucose monitoring at baseline (65% and 72%, respectively) and no statistically significant change over time. In the high engagement group, high engagement with daily weight and glucose monitoring at baseline was observed (82% and 94%, respectively). In this group, a slight but statistically significant increase in weight monitoring was observed over the 6 months, whereas glucose monitoring did not change. The final three-class model was chosen based on a model fit procedure according to AIC and BIC values (Table 4). For all devices, model fit was improved in the three-class model compared to the two-class model. However, for weight and glucose, fit improved only marginally in the four-class model relative to the three-class model. For physical activity, the four-class model could not be produced. The final three-class models were based on weight and glucose device engagement trajectories because the engagement rate for the physical activity device was consistently high over time for all participants with very small variabilities.

Figure 1. Empirical plots (mean, SEM) for biweekly engagement trajectories for each mobile health device over all 6 months.
Table 3. Latent class growth analysis multitrajectory model results for engagement with a wireless weight scale and glucometer (N=60).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Group membership</th>
<th>B</th>
<th>SE</th>
<th>t_1</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>Low Engagement</td>
<td>40%</td>
<td>6.40</td>
<td>6.21</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Medium Engagement</td>
<td>33%</td>
<td>6.30</td>
<td>5.27</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>High Engagement</td>
<td>27%</td>
<td>5.91</td>
<td>4.57</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Glucose</td>
<td>Low Engagement</td>
<td>–0.09</td>
<td>0.02</td>
<td>–3.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Medium Engagement</td>
<td>0.005</td>
<td>0.002</td>
<td>2.82</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td>High Engagement</td>
<td>0.02</td>
<td>0.008</td>
<td>2.34</td>
<td>.02</td>
</tr>
</tbody>
</table>

Figure 2. Empirical summary plot for biweekly engagement trajectories with the (A) glucometer and (B) wireless weight scale by different engagement groups.
Table 4. Model fit by number of latent classes modeleda.

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Weight</th>
<th>Percent per class</th>
<th>Glucose</th>
<th>BIC</th>
<th>Percent per class</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AICb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>–420.1</td>
<td>–427.4</td>
<td>–457.4</td>
<td>–464.7</td>
<td>56.3/43.7</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>–368.4</td>
<td>–379.9</td>
<td>–408.3</td>
<td>–419.8</td>
<td>45.1/25.0/29.8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>–352.2</td>
<td>–368.0</td>
<td>–393.3</td>
<td>–409.0</td>
<td>29.0/16.7/24.4/29.9</td>
<td></td>
</tr>
</tbody>
</table>

aSample size per class is based on most likely class membership.
bAIC: Akaike information criterion
cBIC: Bayesian information criterion.

Associations of Self-Monitoring Behavior Phenotypes With Participant and Clinical Characteristics

The results of bivariate analyses examining how self-monitoring phenotypes are related to participant characteristics are presented in Table 5. Self-monitoring phenotype was significantly associated with age, in that those in the low and waning engagement group were younger compared to those in the medium and consistently high engagement groups. Self-monitoring phenotype was also significantly related to gender as there were less women in the high and medium engagement groups compared to the low engagement group. Race was also significantly associated with group membership, in that those in the low engagement group were more likely to be nonwhite than those in the high engagement group. Insulin dependence was not significantly associated with engagement group.

Table 6 summarizes the association between the self-monitoring behavior phenotypes and HbA1c at baseline, 6 months, and the change between baseline and 6 months. Although the change in HbA1c from baseline to 6 months did not differ according to engagement group, the low engagement group had higher baseline HbA1c values compared to those of the medium and high engagement groups. This trend continued at the 6-month follow up, with higher HbA1c values in the low engagement group compared to the medium and high engagement groups.

Table 5. Bivariate relationships between baseline demographic and clinical characteristics and engagement group membership.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Low Engagement (24/60, 40%)</th>
<th>Medium Engagement (20/60, 33%)</th>
<th>High Engagement (16/60, 27%)</th>
<th>Test statistic</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>49.1 (13.0)</td>
<td>57.9 (10.5)</td>
<td>60.6 (6.3)</td>
<td>F2, 57= 6.48</td>
<td>.003</td>
</tr>
<tr>
<td>Gender: Female, n (%)</td>
<td>20 (47)</td>
<td>10 (23)</td>
<td>13 (30)</td>
<td>χ2=6.96</td>
<td>.03</td>
</tr>
<tr>
<td>Incomea, mean (SD)</td>
<td>4.3 (2.2)</td>
<td>5.7 (2.0)</td>
<td>5.9 (2.3)</td>
<td>F2, 53=3.13</td>
<td>.05</td>
</tr>
<tr>
<td>Race: Black/Non-White, n (%)</td>
<td>20 (51)</td>
<td>11 (28)</td>
<td>8 (21)</td>
<td>χ2=6.01</td>
<td>.05</td>
</tr>
<tr>
<td>Insulin dependent, n (%)</td>
<td>14 (48)</td>
<td>9 (31)</td>
<td>6 (21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (33)</td>
<td>11 (36)</td>
<td>10 (32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>9.01 (2.13)</td>
<td>7.61 (1.17)</td>
<td>7.34 (1.16)</td>
<td>F2,56=6.30</td>
<td>.003</td>
</tr>
</tbody>
</table>

aIncome categories: 1=<$10,000, 2=$10,000-19,999, 3=$20,000-29,999, 4=$30,000-39,999, 5=$40,000-49,999, 6=$50,000-59,999, 7=$60,000-79,999, 8=$80,000.

Table 6. Hemoglobin A1c levels (mean, SD) at baseline, 6 months, and change according to multitrajectory engagement group (N=60).

<table>
<thead>
<tr>
<th>Time point</th>
<th>Low Engagement (n=24)</th>
<th>Medium Engagement (n=20)</th>
<th>High Engagement (n=16)</th>
<th>F df</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>9.01 (2.13)</td>
<td>7.61 (1.17)</td>
<td>7.34 (1.16)</td>
<td>6.30</td>
<td>2.56</td>
</tr>
<tr>
<td>Six months</td>
<td>8.64 (2.54)</td>
<td>7.09 (1.45)</td>
<td>7.16 (1.23)</td>
<td>3.96</td>
<td>2.49</td>
</tr>
<tr>
<td>Change from baseline to 6 months</td>
<td>0.0 (2.23)</td>
<td>-0.44 (1.07)</td>
<td>-0.19 (0.64)</td>
<td>0.38</td>
<td>2.48</td>
</tr>
</tbody>
</table>
Discussion

Consistent self-monitoring and self-management of T2DM improves health outcomes [1,23,24]. Given the growing popularity of using mHealth devices to facilitate self-monitoring, engagement with mHealth devices has become an important tool to develop digital phenotypes based on individuals’ self-monitoring behaviors. This study is among the first to operationalize digital phenotyping of self-monitoring behaviors by applying LCGA modeling on engagement trajectory data from multiple mHealth devices.

Overall, individuals’ engagement with an accelerometer to monitor daily physical activities was consistently high (>82%) for the participants over time. Similar patterns were observed in other mHealth studies [25,26]. This may be due to the passive nature of data collection and transmission of these devices. We were able to identify three distinct digital phenotypes of self-monitoring behaviors using engagement trajectories of wireless weight scales and glucometers. There was a low and waning engagement group (24/60, 40%), a medium engagement over time group (20/60, 33%), and a consistently high engagement group (16/60, 27%). Specifically, the low engagement group started with low engagement in using both the wireless scale (40%) and glucometer (58%), and then the level of engagement rapidly decreased in the first 2 weeks.

These results are similar with those of other studies focused on the use of mHealth technologies or devices for chronic disease management [7]. However, our study provides further evidence by identifying individuals with low engagement in the first couple weeks, demonstrating the need to allocate additional intervention or resources since it is likely that waning engagement will be observed over time. For people who are highly engaged initially, we could consider providing minimum support to save resources as they will be more likely to stay engaged over time.

Our findings also demonstrate how engagement with mHealth devices varies according to patient demographic and clinical characteristics. The individuals in the high and moderate engagement groups were older, included more men, had higher income levels, were more likely to be white, and had better HbA1c values at baseline. By contrast, the low and waning engagement group members were younger, included more women, had lower incomes, were more likely to self-identify as black or nonwhite, and had poorer control of their T2DM. Participants who are insulin-dependent may be required to self-monitor their blood glucose daily or even multiple times a day based on instructions from their primary care physician. This will certainly increase the motivation to engage with the glucometer or even wireless scales for the study participants. However, we did not find any significant association between insulin dependence and engagement group. This implies that we may need to provide further support to this high-risk group. The baseline characteristics of our sample are similar to those of prior research in that lower income individuals, nonwhite individuals, and women face more challenges in controlling glycemia, experience more T2DM complications, and have higher mortality rates [6,27]. We hypothesize that the younger patients in our study may have had lower engagement due to competing demands on their time (eg, caring for family, work), more comorbidities, or having been diagnosed with diabetes at a younger age. Digital phenotypes of self-monitoring behaviors can identify patients who may need the most support in changing health care behaviors and can inform strategies to tailor the use of mHealth tools in the delivery of self-management interventions. This result also indicates that different retention approaches may be needed for certain populations to maintain engagement with mHealth tools in support of T2DM self-management.

As discussed above, individuals with well-controlled baseline HbA1c were more likely to be in the consistently high engagement group. Not surprisingly, these individuals continued to have better controlled HbA1c at the 6-month follow up. However, the change in HbA1c value between baseline and the 6-month follow up did not differ according to different phenotypes of self-monitoring behaviors or engagement groups. This indicates that good self-monitoring behaviors through active engagement with mHealth devices is helpful in maintaining well-controlled HbA1c, but does not necessarily further reduce HbA1c.

Limitations to the study include that the sample was obtained from a single site in the southeastern United States, which may not be representative of all patients with T2DM. A larger-scale study that includes more patients from different regions would yield more generalizable findings to a broader population. Such a study would also help to identify more complex patterns in engagement trajectories and more specific strategies in delivering behavior change interventions. Self-monitoring also occurred for only 6 months, which did not allow for examination of long-term patterns in a complex chronic illness such as T2DM. There are several factors that may affect a patient’s motivation to engage with the device and self-monitoring that was not accounted for in our analysis. First, this was an observational study and the participants were provided with different mHealth devices, which they could keep if they completed the study. Although we did not have any specific requirement or incentive for participants to use the device during the follow-up time, the ability to keep the device may have some implications in retaining their participation in the study. Second, during the 6-month follow-up period of the study, we conducted 20 interviews with the participants to view their data and gain perspectives on using real-time data collections to support self-monitoring. This may have also potentially affected the motivation for participants to engage.

In conclusion, T2DM is a challenging disease that requires frequent self-monitoring and consistent self-management. Digital phenotyping on self-monitoring behaviors using LCGA can help to identify subgroups of individuals with distinct engagement trajectories. Future research should focus on methods to develop tailored mHealth interventions based on the influence of different phenotypes of individuals on their self-monitoring behaviors.
Acknowledgments

This study was funded by a grant from the US National Institutes of Health, the National Institute of Nursing Research (NINR R15NR015890 to RS, QY, DS, AV, MC), a Duke University Center for Nursing Research Pilot grant (QY), and a Duke University Data+ grant (RS). This study was also supported by the Building Interdisciplinary Careers in Women’s Health Career Development Award (K12HD043446 to DS); US Department of Veterans Affairs Office of Academic Affiliations (grant No. TPH 21-000 to AL); Durham Center of Innovation to Accelerate Discovery and Practice Transformation (Grant No. CIN 13-410 to MC and AL); and the US National Institutes of Health (NINR F31NR018100 to JV). The content is solely the responsibility of the authors and does not necessarily reflect the position or policy of Duke University, the US Department of Veterans Affairs, or the US government. The authors acknowledge iHealth for donation of devices and would like to thank Angel Barnes, the study coordinator, for her dedication to assure the accuracy of the data, and Judith C. Hays and Karen Judge for editorial assistance.

Conflicts of Interest

DS is a consultant with Omada Health. The other authors declare no conflict of interest.

References


Abbreviations
AIC: Akaike information criterion
ANOVA: analysis of variance
BIC: Bayesian information criterion
HbA1c: hemoglobin A1c
LCGA: latent class growth analysis
mHealth: mobile health
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 08.01.20; peer-reviewed by H Mehdizadeh, T Powell-Wiley; comments to author 21.03.20; revised version received 30.03.20; accepted 31.03.20; published 11.06.20.

Please cite as:
URL: https://mhealth.jmir.org/2020/6/e17730
doi: 10.2196/17730
PMID: 32525492
User Retention and Engagement With a Mobile App Intervention to Support Self-Management in Australians With Type 1 or Type 2 Diabetes (My Care Hub): Mixed Methods Study

Mary D Adu¹, BSc, MSc; Usman H Malabu¹, MD, FRACP; Aduli EO Malau-Aduli², BSc, MSc, PhD; Aaron Drovandi¹, BPharm, MPharm, PhD; Bunmi S Malau-Aduli¹, BSc, MSc, PhD

¹College of Medicine and Dentistry, James Cook University, Townsville, Australia
²College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Australia

Corresponding Author:
Bunmi S Malau-Aduli, BSc, MSc, PhD
College of Medicine and Dentistry
James Cook University
1 James Cook Drive
Townsville, 4811
Australia
Phone: 61 747814418
Email: bunmi.malauaduli@jcu.edu.au

Abstract

Background: Mobile health apps are commonly used to support diabetes self-management (DSM). However, there is limited research assessing whether such apps are able to meet the basic requirements of retaining and engaging users.

Objective: This study aimed to evaluate participants’ retention and engagement with My Care Hub, a mobile app for DSM.

Methods: The study employed an explanatory mixed methods design. Participants were people with type 1 or type 2 diabetes who used the health app intervention for 3 weeks. Retention was measured by completion of the postintervention survey. Engagement was measured using system log indices and interviews. Retention and system log indices were presented using descriptive statistics. Transcripts were analyzed using content analysis to develop themes interpreted according to the behavioral intervention technology theory.

Results: Of the 50 individuals enrolled, 42 (84%) adhered to the study protocol. System usage data showed multiple and frequent interactions with the app by most of the enrolled participants (42/50, 84%). Two-thirds of participants who inputted data during the first week returned to use the app after week 1 (36/42, 85%) and week 2 (30/42, 71%) of installation. Most daily used features were tracking of blood glucose (BG; 28/42, 68%) and accessing educational information (6/42, 13%). The interview results revealed the app’s potential as a behavior change intervention tool, particularly because it eased participants’ self-care efforts and improved their engagement with DSM activities such as BG monitoring, physical exercise, and healthy eating. Participants suggested additional functionalities such as extended access to historical analytic data, automated data transmission from the BG meter, and periodic update of meals and corresponding nutrients to further enhance engagement with the app.

Conclusions: The findings of this short-term intervention study suggested acceptable levels of participant retention and engagement with My Care Hub, indicating that it may be a promising tool for extending DSM support and education beyond the confines of a physical clinic.

(JMIR Mhealth Uhealth 2020;8(6):e17802) doi:10.2196/17802

KEYWORDS
mobile apps; engagement; retention; diabetes mellitus, self-management; behavioral intervention technology
Introduction

Background
Mobile health (mHealth) apps offer a unique opportunity to deliver health promotion interventions to reach any population due to their ubiquitous nature [1,2], with some developed specifically to support diabetes management [3,4]. However, these mHealth interventions suffer from low participant retention [5,6] and nonusage attrition [6,7]. Therefore, more engaging interventions are required to address these concerns [8,9] through user-centered and iterative approaches that integrate input from users and other relevant stakeholders in app design and development. This approach is necessary to provide interventions that meet user requirements and ensure greater retention, uptake, engagement, and sustainability [9,10].

Retention
Inadequate participant retention is a major methodological challenge experienced by many mHealth app interventions [11]. Low retention rates and lower statistical power threaten outcome validity [6] and serve as a major reason for premature trial termination [12]; hence, pilot studies are important before conducting large-scale studies. The evaluation of participant retention levels enables researchers to assess the relevancy and tendency for sustainable implementation of intervention ideas [13]. In addition, the assessment reveals any required research methodology modification [13] in preparation for future large-scale research.

Engagement
An effective mHealth intervention requires not only retention but also continuous and active engagement by users, as lack of engagement leads to study dropout and dampening of the treatment effect [6,11]. User engagement refers to interaction, experience, perceived usefulness, and desire to use the intervention repeatedly over a long period of time [14,15]. The degree to which users engage with a health app signifies their willingness to invest time, attention, and emotion into the use of the technology to satisfy and eventually achieve their pragmatic needs (such as self-management) [14]. Measurement of users’ engagement can be long or short term in nature with short-term measurement reflecting initial adoption of the intervention and the tendency of apps to successfully engage users in the long term [14]. Although, there are various approaches to measuring engagement with apps, system usage data and user-reported interactions with the system using specific techniques such as questionnaires and interviews are the most relevant in the context of short-term measurement [9,14,16].

System usage is measured through the collection of noninvasive data on the frequency of access to the app, push notifications opened, and average time spent per usage [17,18]. This provides information on user participation with specific target behaviors and frequency of access to the corresponding app features [19,20]. On the other hand, user-reported approaches reveal users’ experiences related to behavioral engagement with the intervention [14,16]. This is necessary to assess intervention tendency to foster achievement of behavioral goals when used over a long period.

Behavioral Engagement Framework
Rate of use alone is not a sufficient indicator of engagement with an mHealth intervention [9]. There must also be an assessment of engagement with the behavioral goal of the intervention to ascertain the intervention’s potential as an effective tool to support behavioral change. One possible way to achieve this is by assessing users’ engagement with the process of achieving behavioral change. Behavioral change is fostered by intervention components that motivate users to achieve a behavioral goal (in this case, diabetes self-management (DSM) behaviors) [9]. Assessing engagement in behavioral change process requires the use of models and frameworks that reveal the relationship between factors in a system for the realization of a defined goal [21].

Within the field of mHealth engagement, models and frameworks provide a richer understanding of the core components that influence user engagement to achieve the behavioral goal of the intervention [22]. The concept of behavioral engagement is complex and includes the extent to which users interact with the intervention. Major considerations include the quality of users’ experience with the technology [23] and if they have engaged with it as needed [7] or as intended [23]. The behavioral intervention technology (BIT) model by Mohr et al [24] describes the full range of components that must be available in a technology to influence engagement with behavioral change and its potential as an effective intervention to attain a behavioral goal.

The BIT framework [24] was utilized in this study as it describes the theoretical components necessary in the conceptualization of mHealth and also instantiates the necessary components for its implementation. The theoretical level covers the overall goal (why) or reason for mHealth development and how specific aims related to the goal could be achieved through the required behavioral change strategies. Each strategy is instantiated by elements: features (what) available in the intervention. In addition, the characteristics (technic) of the intervention affect how an element is displayed to the users as well as their perception about the intervention. Finally, the workflow (pattern of use) describes when and under what conditions BIT interventions will be delivered. Therefore, the BIT model explains that achieving an intervention goal is fostered through relationships between the components of aims, behavioral change strategies, elements, characteristics, and pattern of use of the intervention [24]. We used this model to interpret our qualitative findings, allowing for an open approach to the concept of behavioral engagement, focusing on exploring the tendency of My Care Hub as an intervention tool for diabetes behavioral engagement.

Study Context and Objectives
Owing to poor retention and engagement with previous diabetes apps, we performed an initial study to explore user needs and preferences to foster engagement with a diabetes app [25], which was used to develop a new app called My Care Hub [26]. Patients with diabetes who interacted with a prototype of My...
Care Hub reported that it was easy to use and that the educational contents were valuable in raising awareness about the importance of DSM and increased motivation to engage in self-management activities [26]. Although the usability of the app was satisfactory, it was unclear if My Care Hub has the potential to retain and engage users and if its components meet the requirements of a supporting tool to foster engagement with DSM.

Therefore, this study aimed to examine levels of user retention and engagement with My Care Hub in a short-term single-arm pilot trial. Retention was measured through completion of follow-up surveys, and engagement with the app was assessed in 2 areas: (1) system usage data and (2) qualitative feedback from users on behavioral interactions with the intervention. We expect that the app’s contents and features, which were developed based on results from our previous study on users’ needs [25,26], would result in high participant retention and greater engagement during the short trial period. Understanding these factors is critical in identifying areas where intervention design may need improvement and inform plans for future trials of mHealth interventions such as My Care Hub.

Methods

This study received ethics approval from the James Cook University Human Research Ethics Committee (reference #H7716). Participants were informed about the study aims, and consent was implied by survey submission. Verbal consent was obtained for telephone interviews.

Study Design and Sample Size

This study utilized a sequential explanatory mixed methods design with quantitative surveys and qualitative interviews. This design captures both the engagement with technology and the process of behavioral change by triangulating the results of multiple measures [27]. This provides information about how users react to the contents and design of the intervention and offers an explanation for why users interact with the intervention in a particular way [9]. This study was conducted from August to October 2019, where each participant was given 3-week access to the app. Following this period, participants filled out a survey and were invited to participate in a telephone interview to better understand their interaction with the app.

The study used a maximum variation purposive sampling tailored to recruit participants who showed interest in the study within the time available. This sampling method is appropriate for an implementation feasibility assessment as related to this study [28]. The components of the pilot testing that relate to retention and engagement with the app are presented in this paper.

Recruitment and Eligibility

Participants were recruited through a single invitation email sent to patients registered with the Australian National Diabetes Service Scheme. Email invitations were limited to patients who have type 1 or type 2 diabetes and live in North Queensland, Australia. North Queensland has a relatively high prevalence of diabetes [29] and socioeconomic disadvantage, which can affect accessibility to regular diabetes support services [30]. Therefore, the use of mHealth interventions to provide DSM support may be essential among this population. Other eligibility criteria included ownership of an Android-operating smartphone, having a current recommended blood glucose level (BGL) target of 4 to 10 mmol/L [31], and being aged 18 to 65 years. The upper age limit was chosen because of the less stringent glycemic recommendations for many older adults who are above 65 years. Patients were excluded if pregnant or currently using an app with an educational component to support their diabetes management.

Enrollment and App Orientation

Participants enrolled through the web by completing an eligibility screening form, providing consent, and completing the baseline survey, which entailed questions regarding socio-health demographics, email address, and residential postcode. Participants were emailed a unique code to enable them to download My Care Hub from Google Play store of any android-powered phone, an app manual, and a 5-min video explaining how to install the app, features, and functionalities. Participants could contact the first-named author (MA) for assistance with technical difficulties or for study clarification. It was emphasized that there was no limit to the frequency of use of My Care Hub as participants could engage with it at a level they considered useful and desired. My Care Hub is intended to be a stand-alone intervention; therefore, push notifications (aimed at improving patients’ awareness about diabetes distress and potential ways to reduce its impact on their self-management) were sent from the app during the first 2 weeks of the intervention and withheld in the third week to see the achievable level of engagement with the app with or without push notifications. Throughout the study period, no log-in reminders or calls were made from the study researchers to participants.

Intervention Overview

A detailed description of the development of My Care Hub and the methods of usability studies have been previously published [26]. In brief, the goal of My Care Hub is to provide support and education that facilitates positive behavioral change in diabetes management. The app was specifically designed for type 1 diabetes patients with standard Australian BGL recommendations of 4 to 8 mmol/L for fasting and <10 mmol/L 2-hour postprandial, and for type 2 diabetes patients with recommended fasting BGL of 6 to 8 mmol/L and 2-hour postprandial levels of 6 to 10 mmol/L. The app incorporates multiple functions and features to foster engagement with the app within 3 broad categories: documentation, analytics, and education.

In documentation features, users can manually input data for tracking BGL, physical activity, the carbohydrate content of foods eaten, and body weight. Analytic features provided a graphical output of each documentation feature, thus offering users the ability to visually inspect their logged data over time. Education was provided through 4 main features. First, users can review a variety of actionable textual information related to healthy food choices, self-monitoring of BGL, medication, reducing risk, healthy coping, problem solving, and physical activities. Second, users can look up information related to...
carbohydrate and calorie content of common foods in Australia (categorized under fruits and vegetables, egg and meat, dairy, grain and legumes). Third, the BGL feature provided immediate tailored feedback to every inputted data, driven by a decision-based system. The system is controlled by the value of logged BGL (either within or beyond the standard range), type of diabetes, and the indicated period of BGL measurement (either fasting or 2 hours postprandial). Messages were health-promoting and motivational information aimed at supporting behavioral skills building for self-management practices. Finally, the app provided education through daily push notifications aimed at improving awareness about diabetes distress and encouraging patients to focus on potential ways to reduce its impact on their self-management. Push notifications were terminated at the end of the second week. Sample screenshots are provided in Multimedia Appendix 1.

Postintervention Data Collection
At the end of the study, participants were sent an email (with a reminder email sent to noncompleters), which directed them to the poststudy survey on the acceptability of the app and its preliminary efficacy (results will be reported in future publications). Through this survey, participants were also invited to participate in individual telephone interviews to further understand their perception of the app. Participants who completed the poststudy survey were awarded an electronic gift (e-gift) card worth Aus $40 (US $25.07). All telephone interviewees were contacted within 3 weeks of completing the survey and awarded an additional Aus $20 (US $12.53) e-gift card.

Measures
Retention was assessed using the following indicators of study completion per protocol: number of participants enrolled, number of participants who used the app during the intervention period, and completion rate of the poststudy survey.

Engagement with My Care Hub was measured using participants’ app usage log and verbal feedback. App usage data were extracted from the app’s activity database. The following time frames were considered: (1) date of log-in into the app to 2 weeks of use when the daily push notification was administered (referred to as week 1 and week 2) and (2) data during the third week (referred to as week 3) after the termination of push notifications. Key metrics collected from the database included app use (number of active users, frequency of daily access to app), data logs/time spent (for BGL, exercise, food activity, and weight), and number of opened notifications. Metrics were presented using an adapted version of the Frequency, Intensity, Time, and Type (FITT) principle index [32,33]. This index explores multidimensional domains of usage data, which provides greater insight into interaction with an mHealth app. Event count in the app was available for active (documentation features) and passive (viewing of educational screen) app features. Data had to be logged/saved in the documentation features before it could be counted as an active event as the app discarded data not logged after 30 min of inactivity. Users had to exit from an educational screen before it was counted as a passive event, and no maximum count per user was stipulated. This implies that the total count of passive events could be higher if a screen was viewed more than once. The FITT index used in this study is as follows:

1. Frequency index ($F$): This subindex is an attention proxy that provides information on how often a participant uses the app. It recognizes the number of users who return to use the app and active app users in each time period.
2. Intensity index ($I$): This subindex denotes the proportion of users who interact with each feature in the app. In total, 2 metrics were used in the assessment of $I$. These are the frequency of daily use of app features ($I_{da}$) and number of push notifications opened versus the total sent ($n=14$) in 2 weeks ($I_{dp}$). In addition, intensity also measures the proportion of app features used out of the total available features.
3. Type index ($T$): This provides information on the form of engagement based on actions performed by users using the available app features. In this study, the type of action was categorized as active denoted as $T_{sa}$ (use of documentation features for self-monitoring), and passive ($T_{sp}$, reading information on educational contents in the app).
4. Time index ($T_{i}$): This measures the duration of engagement, which signifies attention to the app as a function of daily event duration with each app feature.

Interviews
Interviews were conducted using a semistructured interview guide that explored behavioral engagement with the app through questions on patterns of use, perceived ease of use, perceived usefulness of app features enabling motivation for continued engagement with DSM, and recommendations on how the app could be improved. The interview guide has been provided in Multimedia Appendix 2.

Interviews were conducted by 1 author (AD), who is well experienced in qualitative research. The interview guide was pilot tested between MA and AD before actual use. The interviewer was located in a private office at James Cook University, Australia, while participants were asked if they were in a comfortable location before commencement of the interview. The first 3 interviews were used to reflect on the guide, although there were no resultant changes. Data saturation was achieved as judged by no emerging new information [34] after completing the 15th (of 17) interview. Interviews were audio recorded, and none of the participants had a previous relationship with any of the authors.

Data Analysis
Descriptive statistics were calculated for all quantitative variables. Baseline characteristics comparison between those who completed the study and those who did not were done using a Pearson chi-square test. All statistical analyses were performed using SPSS version 23 [35].

Interviews were completed in an average of 15 min (range 9-30 min). Participant responses were transcribed verbatim by 1 researcher (AD). In this analysis, a combination of data and a concept-driven strategy was applied. Initially, inspired by the work of Schreier [36], 2 researchers (MA and AD) independently used a data-driven strategy to obtain an overview
of the data, and then similar text segments were selected and sorted using coding. Coded segments were grouped to identify recurring themes from the data. Themes were compared between the 2 authors, discussed with a third author (BM), adjusted, and an agreement was reached about the main themes. Subsequently, the authors analyzed the themes by applying a concept-driven strategy in accordance with the BIT framework [24] to assess behavioral engagement with the intervention. We identified and described the BIT components in the My Care Hub app that could potentially enhance behavioral engagement with it. These components overlap and diverge within the identified themes, which are presented using representative quotes affixed with an assigned number code and the type of diabetes the respondent has (for instance, respondent 3 with type 1 diabetes; P003, T1D and respondent 4 with type 2 diabetes; P004, T2D). The conduct and reporting of the interviews followed the consolidated criteria for reporting qualitative research (Multimedia Appendix 3) [37].

Results

Participant Characteristics

Participant demographics and health characteristics are shown in Table 1. Participants were predominantly male (31/50, 62%), had type 2 diabetes (36/50, 76%), and aged between 20 and 64 years (mean 49.12, SD 12.34 years). On average, the recommended BGL in enrollees was as follows: for fasting, 4.58 (SD 0.78; range 4-6 mmol/L), and for 2-hour postprandial, 7.01 (SD 1.02; range 6-10 mmol/L). Most participants were diagnosed as having diabetes in the last 5 years (27/50, 54%), and an equal proportion rated their health status as being fair or good (20/50, 40%). Most had a technical college education or higher (39/50, 78%) and were employed (31/50, 62%). Only a few had previously used a health app to manage diabetes in the past (16/50, 32%). The linking of participants’ postcode to the Australian Standard Geographical Classification System [38] indicates the geographic location of the majority to be rural (37/50, 74%).

Of the 22 participants who indicated an interest in participating in the interview, only 17 were contactable within 3 call attempts. Most were males (12/17, 71%), had type 2 diabetes (13/17, 77%), and had been diagnosed for an average of 6 years (range 1-17 years). Overall, participants were between the ages of 36 to 64 years (mean 51.58, SD 11.31), except for one who was aged 20 years.

Retention

Of the 4984 patients who were emailed an invitation to participate in the study, 79 (1.59%) completed the eligibility form. However, only 84% (67/79) of those who responded met the inclusion criteria and were provided access to download the app. Some participants (17/67, 25% of those eligible) failed to log in to the app, resulting in 50 enrolled participants (75% of eligible participants). Most enrollees (43/50, 86%) activated the app within the same day (range 0-5 days) of having access to it. One participant logged out of the app on the second day of installation stating that it did not meet her requirement. At the end of the study period, 41 of the enrolled participants completed the study per protocol by providing feedback about the app using the poststudy survey (retention rate: 41/50, 82%). Reasons for noncompletion of the study protocol were not recorded. In assessing baseline characteristics associated with retention, only employment status emerged as a significant predictor, with those unemployed being less likely to complete the study than those who were employed (50.0% versus 14.7%, respectively; \( P = .02 \)). The full details of the demographic variables and comparison between those who completed the study and those who did not are shown in Table 1.
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline (N=50)</th>
<th>Completers (n=41, n (%))</th>
<th>Lost to follow-up (n=9, n (%))</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>31</td>
<td>25 (81)</td>
<td>6 (19)</td>
<td>.75</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td>16 (84)</td>
<td>3 (16)</td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>N/A^a</td>
<td>49.29 (12.74)</td>
<td>48.67 (11.25)</td>
<td>.82</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td>.82</td>
</tr>
<tr>
<td>18-29</td>
<td>5</td>
<td>4 (80)</td>
<td>1 (20)</td>
<td></td>
</tr>
<tr>
<td>30-39</td>
<td>6</td>
<td>5 (83)</td>
<td>1 (17)</td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>12</td>
<td>10 (83)</td>
<td>2 (17)</td>
<td></td>
</tr>
<tr>
<td>50-59</td>
<td>15</td>
<td>11 (73)</td>
<td>4 (27)</td>
<td></td>
</tr>
<tr>
<td>60-65</td>
<td>12</td>
<td>11 (92)</td>
<td>1 (8)</td>
<td></td>
</tr>
<tr>
<td>Type of diabetes, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.81</td>
</tr>
<tr>
<td>Type 1</td>
<td>15</td>
<td>12 (80)</td>
<td>3 (20)</td>
<td></td>
</tr>
<tr>
<td>Type 2</td>
<td>35</td>
<td>29 (83)</td>
<td>6 (17)</td>
<td></td>
</tr>
<tr>
<td>Type 2 medications or not, n (%)^b</td>
<td></td>
<td></td>
<td></td>
<td>.32</td>
</tr>
<tr>
<td>None</td>
<td>2</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td></td>
</tr>
<tr>
<td>Oral drugs alone</td>
<td>33</td>
<td>28 (85)</td>
<td>5 (15)</td>
<td></td>
</tr>
<tr>
<td>Oral and insulin</td>
<td>1</td>
<td>1 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Duration of diagnosis (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
</tr>
<tr>
<td>&lt;5</td>
<td>27</td>
<td>23 (85)</td>
<td>4 (15)</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>10</td>
<td>8 (80)</td>
<td>2 (20)</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td>9</td>
<td>6 (67)</td>
<td>3 (33)</td>
<td></td>
</tr>
<tr>
<td>&gt;16</td>
<td>4</td>
<td>4 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.59</td>
</tr>
<tr>
<td>High school equivalent</td>
<td>17</td>
<td>12 (71)</td>
<td>5 (29)</td>
<td></td>
</tr>
<tr>
<td>Technical college</td>
<td>10</td>
<td>9 (90)</td>
<td>1 (10)</td>
<td></td>
</tr>
<tr>
<td>First degree</td>
<td>11</td>
<td>10 (91)</td>
<td>1 (9)</td>
<td></td>
</tr>
<tr>
<td>Postgraduate</td>
<td>8</td>
<td>7 (88)</td>
<td>1 (12)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>3 (75)</td>
<td>1 (25)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.87</td>
</tr>
<tr>
<td>Caucasian/white</td>
<td>47</td>
<td>38 (81)</td>
<td>9 (19)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>3 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Employment, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.02^c</td>
</tr>
<tr>
<td>Unemployed</td>
<td>8</td>
<td>4 (50)</td>
<td>4 (50)</td>
<td></td>
</tr>
<tr>
<td>Partly/fully employed</td>
<td>34</td>
<td>29 (85)</td>
<td>5 (15)</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>8</td>
<td>8 (100)</td>
<td>0 (0.00)</td>
<td></td>
</tr>
<tr>
<td>Living environment, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.26</td>
</tr>
<tr>
<td>Remote</td>
<td>13</td>
<td>12 (92)</td>
<td>1 (8)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>37</td>
<td>29 (78)</td>
<td>8 (22)</td>
<td></td>
</tr>
<tr>
<td>Usage of smartphone (years), n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.42</td>
</tr>
<tr>
<td>1-5</td>
<td>13</td>
<td>11 (85)</td>
<td>2 (15)</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>28</td>
<td>24 (86)</td>
<td>4 (14)</td>
<td></td>
</tr>
</tbody>
</table>
**Baseline (N=50) vs Completers (n=41), n (%)**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Baseline</th>
<th>Completers</th>
<th>Lost to follow-up (n=9), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;10</td>
<td>9</td>
<td>6 (67)</td>
<td>3 (33)</td>
<td>.93</td>
</tr>
<tr>
<td>Previous use of health apps to manage diabetes, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>16</td>
<td>13 (81)</td>
<td>3 (19)</td>
<td>.38</td>
</tr>
<tr>
<td>Never</td>
<td>34</td>
<td>28 (82)</td>
<td>6 (18)</td>
<td></td>
</tr>
<tr>
<td>Rating of health status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.38</td>
</tr>
<tr>
<td>Poor</td>
<td>1</td>
<td>1 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td>19</td>
<td>14 (74)</td>
<td>5 (26)</td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td>21</td>
<td>17 (81)</td>
<td>4 (19)</td>
<td></td>
</tr>
<tr>
<td>Very good</td>
<td>9</td>
<td>9 (100)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

aN/A: not applicable.

bN=35.

cP<.05.

**App Engagement**

Most (42/50, 84%) enrolled participants logged data into the app at least once (during week 1 of installation) with the frequency index showing that they actively used the app on an average of 11 of the 14 days in the first 2 weeks when push notifications were sent (range 2-14 days; week 1 average: 5.2 days, week 2 average: 4.8 days). This reduced to an average of 4 of 7 days (range 2-5) in week 3: average 3.8 days. Furthermore, all participants who logged in to the app used it during week 1, and most returned to use the app after week 1 (36/42, 85%) and week 2 (30/42, 71%) of installation. With regard to the intensity index related to daily use of each app feature ($I_i$), most participants used features for tracking their BGL (28/42, 68%) and accessing educational information (6/42, 13%) more frequently. The feature with the least daily use was tracking the carbohydrate content of foods (2/42, 2%). All 14 push notification messages during the first 2 weeks (1 per day) sent were published, and on average, 57% (24/42) of participants opened this notification within 24 hours, after which they were automatically deleted. None of the app features were unused. The type index ($T_y$) shows active and passive actions with the My Care Hub. The average frequency of BGL data log per participant in week 1 was 10.85 (SD 9.32; range 1-36), which reduced to 6.75 (SD 7.75; range 1-24) in week 2 and 5.67 (SD 6.05; range 0-22) in week 3. Physical activity logs showed a mean of 4.48 (SD 3.64; range 0-15) in week 1 compared with 2.97 (SD 2.93; range 0-11) in week 2 and 1.69 (SD 1.70; range 0-7) in week 3. Average passive engagement per participant on occasions of viewing screens alone in week 1 was 26.5 (SD 25.1; range: 9-32), 17.55 (SD 7.39; range 7-26) in week 2, and 14.4 (SD 6.13; range 6-24) in week 3. The time index ($T_t$) revealed that, for all events of participants’ visit to the app, an average daily time of 3.56 min (range 1.37-7.48 min) was spent. More time was spent on BGL activity (2.2 min) and accessing the educational tips embedded in the app (1.35 min). Table 2 summarizes the app functions and features, their purposes, usage, and engagement.
Table 2. My Care Hub sections and engagement indices (N=42).

<table>
<thead>
<tr>
<th>Functions and features</th>
<th>Elements</th>
<th>Purpose</th>
<th>User engagement</th>
<th>Average time spent per user per day ($T_i^b$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Documentation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BG activity ($T_{a1}^d$)</td>
<td>BG log</td>
<td>Monitoring and tracking of BG values over time</td>
<td>29 (69)</td>
<td>2 min 2 seconds</td>
</tr>
<tr>
<td></td>
<td>Type of BG</td>
<td>Gain knowledge to support self-management practices</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automatic feedback (as part of education)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity ($T_{a2}^c$)</td>
<td>Log of time spent on physical activity</td>
<td>Monitoring of physical activity behavior over time</td>
<td>4 (10)</td>
<td>0 min 7 seconds</td>
</tr>
<tr>
<td></td>
<td>Calories used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Place</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Food activity ($T_{a3}^a$)</td>
<td>Record of food intake</td>
<td>Monitoring and tracking of food intake and their carbohydrate content over time</td>
<td>1 (2)</td>
<td>0 min 17 seconds</td>
</tr>
<tr>
<td></td>
<td>Log of carbohydrate content of food</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weight log ($T_{a4}^d$)</td>
<td>Body weight log</td>
<td>Body weight assessment over time</td>
<td>2 (5)</td>
<td>0 min 22 seconds</td>
</tr>
<tr>
<td>Analytics ($T_{a5}^e$)</td>
<td>Graphical display of data log into each documentation feature</td>
<td>Keeping track of trends in lifestyle activities and observe impact on BGL$_f$ over time</td>
<td>3 (6)</td>
<td>0 min 20 seconds</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Textual screens for management tips and food choices ($T_{eb}^c$)</td>
<td>Information on behaviors in DM$_g$ management</td>
<td>Assess current knowledge on DSM$_h$</td>
<td>6 (13)</td>
<td>1 min 35 seconds</td>
</tr>
<tr>
<td></td>
<td>Information on average carb and calorie content of common Australian foods</td>
<td>Review carbohydrate content of foods to make healthy choices.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Push notifications ($T_{eb}^d$) and ($I_{i2}^a$)</td>
<td>Messages on diabetes distress</td>
<td>Create awareness about diabetes distress and ways to reduce its impact on self-management</td>
<td>24 (57)</td>
<td>—j</td>
</tr>
</tbody>
</table>

$a_{i, f}$ intensity index for frequency of daily use.

$T_i$: time index.

$^b$ $T_i$: time index.

$^{BG}$: blood glucose.

$^{d}T_{a1}$: type index for active app use.

$^{c}T_{a2}$: type index for passive app use.

$^{f}$ BGL: blood glucose level.

$^{g}$DM: diabetes management

$^{h}$DSM: diabetes self-management.

$^{i}I_{i2}$: intensity index for number of push notifications opened.

$^{j}$Not captured due to the tracking limitations of the system usage database.

**Interview Results**

Different themes emerged from the data with interconnection among the themes over the course of My Care Hub usage. Overall, the results suggest that the use of the app has the potential to ease the effort in aiming for improved self-management and for better awareness of BGLs. In addition, participants provided their recommendations for extra functionalities that may further enhance engagement with self-management behaviors. We present our findings in relation to themes related to components of the behavioral intervention model [24] used for this study, which are outlined in Table 3.
### Table 3. Summary of behavioral intervention technology model as adapted to My Care Hub intervention.

<table>
<thead>
<tr>
<th>BIT(^a) components</th>
<th>BIT(^a) components</th>
<th>Details in MCH(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why</td>
<td>Broader goal: self-management support</td>
<td>Aims:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Improved BG(^c)—long-term impact</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Increased physical activity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Healthy eating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Decreased diabetes stress</td>
</tr>
<tr>
<td>How</td>
<td>Behavioral change strategies</td>
<td>• Elements or strategies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Documentation and Analytics:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accountability; Clarity of self-management activities and impact;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improved awareness of BG(^c) levels; Mindfulness of calorie consumption</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Feedback response:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reinforced recommendation of HP(^d); Informative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Carbohydrates in foods:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Guidance on meal planning; Knowledge provision and reinforcement</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Educational tips:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knowledge reinforcement</td>
</tr>
<tr>
<td><strong>Instantiation</strong></td>
<td>Documentation (logs) and analytics:</td>
<td></td>
</tr>
<tr>
<td>What</td>
<td>Elements (app features)</td>
<td>• Feedback response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Carbohydrates in foods</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Educational tips screen</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Push notifications</td>
</tr>
<tr>
<td>How (technic)</td>
<td>Characteristics</td>
<td>Aesthetic:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Beautiful</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ease of use:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Simple and straight forward</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Few difficulties</td>
</tr>
<tr>
<td>When</td>
<td>Pattern of use</td>
<td>User defined</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Type of diabetes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Established self-management routines</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Daily Partly, with reasons</td>
</tr>
</tbody>
</table>

\(^a\)BIT: behavioral intervention technology.  
\(^b\)MCH: My Care Hub.  
\(^c\)BG: blood glucose.  
\(^d\)HP: health provider.

**Pattern of Use (When)**

**User Defined**

Patterns of app use depended on users’ type of diabetes and self-management routines, with most participants using the app multiple times per day, where those with type 1 diabetes input their BGL any time it was measured:

\[ \text{I use it multiple times per day, basically any BGL I took I enter it at any time I took it.} \quad \text{[P001, T1D]} \]

In contrast, participants with type 2 diabetes described that the frequency of usage depends on the self-management activity carried out on that day:

\[ \text{I used it at least once a day. If I had done exercise, then I was putting in an exercise and blood test virtually every day. On every second day I was using it to stick in weight but the exercise was done at a different time.} \quad \text{[P005, T2D]} \]
Conversely, some participants were only able to use the app infrequently because of issues such as limited internet access or multiple competing interests:

> I didn’t use it fully, because at the moment I am having a problem with my internet, so I didn’t get a chance to watch the video that comes with it. [P012,T2D]

> I used it a few times to start with, but then I stopped pretty much because I was juggling between doing a lot of writing, doing a course, and was having other things to do. [P003,T1D]

**Characteristics of the App (How)**

**Simple and Straightforward**

Participants described the design as:

> Very well crafted and well put together, really easy to use [P007, T2D],

and could be used even by the elderly who may not be too proficient in using mobile technology:

> I would even say that like an older person in their 60s or so, once they get an idea of how to use it properly, would have no worries using it if they were in that way inclined. [P012,T2D]

**App Difficulties**

Some participants found a few aspects of the app difficult:

> There was one for the activities you had to put in what calories you might have burned off and I didn’t have a clue how I was going to find out that information. [P013,T2D]

> I had a problem figuring out how to put dates in it, but I think it does it itself, so yeah. [P008,T2D]

**Goal (Why), Elements (App Features; What), and Behavioral Change Strategies (How)**

The goal of developing My Care Hub was to enhance engagement with self-management activities such as improved BG, increased participation in physical exercise, and healthy eating. Participants identified multiple elements (features) that support this overall goal. They also described the perceived benefits (mechanism of action) of each of the elements that encouraged their interaction with it, and toward achieving an improved DSM. The commonly mentioned features are noted below, as well as reasons why participants found the features engaging.

**Documentation/Analytics**

**Accountability**

Participants mentioned that the documentation element strengthened the sense of responsibility to keep up with routines in DSM:

> I liked the activity log, because it gives you accountability, when did you go to the gym, how long were you there, what did you do. [P014,T2D]

**Clarity of Self-Management Activities and Impact**

Participants explained that visualization of logged data using analytics encouraged their interaction with My Care Hub. They noted that the feature provides better clarity on their level of self-care:

> Just the tracking of my fitness, exercise and my blood sugars, it is much better for me seeing it in a graph, makes it really clear how you are going. [P006,T2D]

The feature also hinted at some participants to consult their physician for medication review or consultation if their BGLs were not in the recommended range:

> I liked the graphs…, that was what gave me the red flag…maybe I have to see the doctor to have my medications changed. [P010,T2D]

**Improved Awareness of Blood Glucose Levels**

Participants noted that although they have a BG meter that provides BG measurement history, having the graphical output of their BGL in My Care Hub further improved awareness of any fluctuations in BGLs:

> It was quite good to see longitudinal things, obviously on my blood monitor I can see by just hitting the back key what the previous readings are…. But to see it in a graphical linear form was really good. It showed me where my blood sugar was, if I went up and down. [P005,T2D]

**Mindfulness of Calorie Consumption**

The analytic feature enabled participants to pay attention to daily calorie intake or carbohydrates consumed:

> I liked that idea of putting it all in and seeing how your graphs went up and down, and it sort of kept you a bit more mindful of how many calories or carbs you are eating during the day. [P013,T2D]

**Feedback Response**

**Reinforced Health Provider’s Recommendation**

Feedback received in response to logged BGL is an element that reinforced the doctor’s recommendation about participants’ BGLs. A participant with hypoglycemia unawareness noted that his doctor suggested continuing using the app to serve as an alert in the event of low BGL:

> It is one thing that made me maintain my BGLs. I tend to be what my doctor calls hypoglycaemia insensitive. So, he suggested that I stick with the app because it reminds me to do regular BGL tests to make sure that I am not dropping too low. [P007,T2D]

**Informative**

Feedback feature serves as an alert about a potential problem in users’ BGLs:

> I got confirmation that was somewhat reassuring, I mean if it was out and higher, it just alerts you to a potential problem that you may or may not be aware of. [P005,T2D]

It aided decision making for improved self-management:
If my levels were over the target range, it gave me very helpful ways to reduce the blood glucose level back into the range. [P007,T2D]

Carbohydrate Components in Foods

Guidance on Meal Planning
Participants valued the carbs in foods feature as it provided information about the average carbohydrate and calorie contents of foods. Participants perceived they were better supported in their choice of appropriate foods to eat and avoid exceeding their recommended daily amount of carbohydrate intake. It also provided guidance on food planning:

I try to stay between 20 and 50 grams a day, so the carb counting feature was very useful because then you can make an informed decision on what you are going to put on your plate, and you can plan out your week. [P009,T2D]

Knowledge Provision and Reinforcement
Participants who had difficulties knowing the carbohydrate content of foods found this feature useful through outlining the best foods for consumption to ensure proper health management:

I have a lot of trouble with how much carbohydrate is in one food but it (app) sort of gets you to realise okay then I have got to check on that. [P004,T2D]

Furthermore, engaging with the carbs in foods feature reinforced knowledge and served as a reminder about carbohydrate content in foods:

There is so much to take in, like reading labels, it is so much to take in. So I found it (app) quite interesting that it is a bit more set out with carbs and how much is in it, and some of them are low and you thought it would be high. Just reinforcing the information because I just can’t remember everything. [P014,T2D]

Educational Tips

Knowledge Reinforcement
Educational tips were also acknowledged as a tool for knowledge reinforcement and fostered the use of the app. Participants found information on 7 essential ways to manage diabetes quite useful and reflective:

It is useful, I have got a couple of books, and there is a lot of information, and whilst I may have read it, I am not sure I can regurgitate it. [P005,T2D]

It was just interesting to read it and think about it. [P014,T2D]

In addition, participants felt that the element provided more comprehensive information in comparison with the feedback element:

That (educational tips) was more useful than the little hint things (feedback messages) yeah... I think it probably covered it (all information) fairly thoroughly. [P006,T2D]

Recommendations to Further Improve Engagement With the App
Participants’ recommendations were primarily based on extended functionality in the app, including the following:

1. Automation of data input: Some participants found the manual recording of BGL, physical activity, and carbohydrate content of foods consumed as burdensome and expressed that the addition of Bluetooth, which could automatically extract data from the BGL meter, would not only encourage users’ engagement with My Care Hub but also improve BGL monitoring. Furthermore, the desire for the app to automate the tracking of time spent on physical activities and equivalent calorie expended was expressed. In addition, it was recommended that the app should have features to calculate the calorie content of composite dishes.

2. More analytic histories: Participants suggested extended historical data access and believed this would provide further opportunity to study patterns in self-management activities and have long-term data that could be reviewed by their health care providers.

3. Information update: It was suggested that the Carbs in Foods feature needed more food lists and varieties of composite dishes. Participants suggested that this information could be provided in monthly updates because users’ awareness of finding new information in the app on a regular basis could foster fresh interest in using the app.

4. Feedback on physical activities: The idea of providing motivational feedback in the app, especially when users achieve certain levels of physical exercise, was raised. This behavioral change strategy in My Care Hub is presently limited to the BGL documentation; presumably, participants want an extension of it to the physical activity documentation.

Discussion

Overview
The My Care Hub mobile app intervention was intended to encourage ongoing participation in DSM activities. This paper reports the levels of participant retention and engagement (usage and behavioral aspects) with the technology over a 3-week pilot study. The findings of the study revealed an acceptable level of participant retention with the intervention, where the majority completed the study per protocol. Furthermore, participants reported that the intervention eased and improved their effort in participating in self-management activities. Thus, suggesting the app’s potential as a tool for DSM support and education. Nevertheless, a larger sample and longer-term studies are required to establish these claims.

Participant Retention
The retention rate was relatively high, with more than three-quarters (82%) of participants completing the study per protocol, which is similar to previous short-term pilot studies of diabetes app interventions [39,40]. This indicates that participants were highly motivated and willing to participate in their self-management activities. However, some other pilot studies on DSM support programs reported higher retention.
than this study. For example, Dick et al [41] reported 0% attrition over 4 weeks, whereas Kim et al [42] reported only 3% loss to follow-up over a 3-month pilot testing. Such findings are expected because the studies [41,42] were conducted in controlled settings where participants’ recruitment took place in health care facilities, whereas our study utilized web-based recruitment. Participants are likely to be more committed to the studies when recruited from their care facility and with the knowledge of their care physician [43]. In contrast, studies such as ours that recruited participants through the web may experience a quick loss to follow-up due to a less structured environment [6,11]. Future studies with My Care Hub might consider recruitment from a structured setting as a further strategy to improve participants’ retention. Retention was not influenced by participant characteristics measured, with the exception that unemployed participants were less likely to complete the study, which was contrary to the results of a previous mHealth study [44]. Reasons for this discrepancy are unclear, although despite this difference, 50% of unemployed participants were retained in this study, which is relatively high for web-based interventions. Future research with My Care Hub will explore reasons for higher attrition among unemployed participants and the use of empirical strategies to improve their retention rates.

**Intervention Engagement**

Users in our study actively used the app for 11 of 14 days (11/14, 79%) in the first 2 weeks, where they all used the app at least once during the first week and 85% returned to use the app during week 2 and 71% during week 3. To put these rates into perspective, we refer to studies of Faridi et al [45] and Kim et al [42], who found that 53% and 38%, respectively, of participants used the app for a portion of the 12 weeks intervention duration, where in some cases, there was up to 33% of completely inactive participants [45]. In comparison, our app frequency usage rate can be interpreted as reasonable. However, mobile-based interventions differ widely in terms of population, features, settings, and techniques used to foster engagement. For example, although our intervention was self-directed, and we did not utilize reminders for self-management or data entry, the above-mentioned studies used face-to-face intervention orientation [42,45], automated reminders for diabetes management [45], and physician review of adherence [42]. These disparities may have been a major influence on usage, making direct comparison with other app-based interventions difficult. However, the sharp reduction in app usage during week 3, where only 71% returned to use the app without the push notifications reveals the role of push notification as a feature that could further stimulate users’ engagement with apps [46], especially those with content containing insights into how to overcome barriers to achieving health goals [47] as provided in this study. Nevertheless, some users find push notifications intrusive and annoying, especially when too frequent, thus limiting engagement with the intervention [48]. Hence, health apps should be built in ways that patients can customize and review when they see notifications or adjust the timing to suit the selected period of specific self-management tasks such as physical exercise or BG monitoring.

The intensity of usage showed that participants interacted more with features for monitoring of BGL and physical activities, which are in congruence with previous studies [5,49]. This was confirmed in the interviews where participants mentioned that these documentation features improved accountability for their self-management activities. This may be due to patients’ understanding of the importance of these self-management activities for optimal health outcomes. Another explanation might be because the documentation features were accompanied by analytics that foster improved awareness of BGLs, accountability, and better clarity of self-management activities, as mentioned in the interview. These behavioral strategies in the documentation and analytic features might have encouraged personal reflection among participants, hence the increased intensity of usage.

The active time spent on the documentation features demonstrated that the duration of app usage necessary to generate consistency is a parameter that depends on individual users [50]. This was reflected in the interviews where the pattern of use was denoted by users’ decision on sequence and DSM routine. This result reveals the advantage of a multicomponent intervention such as My Care Hub, which offers users the opportunity to embrace it in ways most relevant to their needs [51]. A user can bypass a feature that they feel does not apply to them, potentially increasing engagement with more relevant areas in relation to their needs. Therefore, the diverse elements available in My Care Hub represent an advancement over many existing diabetes app interventions that consist of only a single element that requires participants to complete a predefined behavioral program [52].

Although the My Care Hub system log recorded participants’ passive usage of the education textual screens, there are no standard measures to compare these data with similar diabetes-focused interventions. However, the interviews indicated that participants appreciated this feature as an important element that provided knowledge reinforcement as a behavioral strategy for DSM. Nonetheless, the app system was unable to capture whether participants were actually reading and comprehending the embedded information or simply clicking them. An approach to address this limitation is to incorporate eye-tracking technology [53] or tailored quizzes [54] into My Care Hub to measure cognitive responses and knowledge acquired through engagement with each information screen. These measures would need to determine if success or failure of a user to acquire knowledge is due to the intervention component delivery mode, users’ engagement with the information, or some other intrinsic factors exclusive to the user.

Generally, engagement indices were initially high but decreased in subsequent weeks. Previous studies using mHealth interventions over short- and long-term periods have identified similar trends [52,55]. This finding was expected, as this study was a real-life pragmatic pilot testing of an app, prone to nonuse or infrequent use because users prefer to engage with apps periodically [55]. In addition, nonusage attrition with mHealth could be due to other reasons such as lack of self-motivation or commitment to change health behaviors [55] and satisfactory attainment of knowledge or skills in managing the disease [52].
Participants’ perceptions related to behavioral change strategies in My Care Hub derived from the documentation, feedback response, calories in foods, and education tips features are consistent with the needs analysis study conducted as part of the predevelopment phase of the app [25]. Both type 1 and type 2 diabetes patients expressed a strong interest in these elements because of their ability to not only foster engagement with an app but also provide benefits for self-management behaviors. This reinforces the notion that benefits derived from an intervention strongly affect users’ experience and, hence, engagement with the technology [23]. As these elements are targeted toward self-monitoring of behavioral activities and the provision of educational information to support those activities, the perceived behavioral change strategies may be an indicator that the app has the tendency to support users to achieve their behavioral goals. Nevertheless, further long-term studies are required to establish this claim.

Perceived ease of use of mHealth positively affects continuance in intention to use [56]. The presentation and characteristics of a technology determine the way users can optimize the elements to achieve their aim and overall behavioral goal [24]. If users enjoy their experience in a digital behavioral intervention, exposure to the behavioral change component will be improved and may subsequently influence behaviors [22]. These were reflected in our study as participants expressed their opinion about the simplicity of My Care Hub and perceived it as uncomplicated and effortless to use. Even when engagement is a purposeful choice and evolves from how people choose to obtain value from their experience, it has to be enabled by the technology and, thus, impacts long-term interaction with such technology [14].

The educational component of the app was informed by our previous study, which shows that information on basic guidelines for the management of diabetes and approaches to problem solving in diabetes were highly desired by both type 1 and type 2 diabetes patients [25]. However, once that knowledge is obtained, there is a tendency for a drop in participants’ rate of use of the app [57]. This highlights that apart from developing an app to meet end-user requirements and perceived relevance to diabetes management, mHealth developers need to consider ongoing novel strategies that will keep participants engaged. Novelty is also a main contributor to app engagement because it prevents boredom [23,58]. The downward trend in engagement indices may be explained by a lack of novelty in the app throughout the study period. Hence, future long-term research with My Care Hub must consider ongoing novel strategies that will keep participants engaged. Such strategies may be achieved by considering the suggestions raised by participants in this study. These include periodic information updates on meals and their corresponding nutrient values. Other suggestions on extended functionality in accessing more historical data, automated data transmission, and feedback on physical activity performance are also potential future improvements of My Care Hub, as they have been proven to have an effect on behavior [58].

Strengths and Limitations

A mixed methods study design was used to evaluate patient engagement with My Care Hub, which is a strength of the study compared with previous studies that have arbitrarily classified engagement as high or low based on frequency of use [52] or overall adherence to the intervention [59]. The unique contribution of this paper is threefold. First, retention with My Care Hub indicates its potential as a relevant behavior change intervention tool for patients with diabetes in rural or remote environments with poorer access to specialist health care services. Second, participants’ engagement based on interaction with multiple intervention elements was measured using the FITT metrics. The use of this measure reveals the level of user engagement with each intervention feature, thus providing results that are beneficial to inform future enhancements of My Care Hub. Although FITT is commonly used in physical activity research [33], to the best of our knowledge, this is the first study to use this measure to assess users’ engagement with a multi-component DSM app. Adjusting the index to measure engagement with the intervention in this study was possible because behavior metrics and physical activities were measured. The use of FITT as a measure provided results that could broadly serve as a reference to evaluate other diabetes mHealth interventions before the execution of a full-scale trial. Third, due to the short intervention period of this study, we employed a theoretical and conceptual framework to confirm the components of BIT present in My Care Hub, as an analog to measures of behavioral engagement with the app. Therefore, the framework served as a predictive device to evaluate the app’s suitability as a behavior change intervention tool. This approach supports a more comprehensive assessment of engagement than most existing short-term pilot studies, which lack theoretical foundations. The use of this framework provides guidance on aspects of mHealth interventions to ensure the development of a meaningful tool that could improve patient engagement with healthy behaviors [24].

This study has some limitations that should be taken into account when interpreting the findings. The short intervention period is acknowledged. However, 3 weeks is the minimum time required for anyone to form a behavioral habit [60], and multiple components as found in our intervention are potentially effective techniques to achieve behavior change [61]. Furthermore, participant recruitment was restricted to a single source, and the sample size was small, thus limiting the sample diversity and generalizability of the results. In addition, the requirement that eligibility includes access to both an Android smartphone and an active email account may imply that the findings may not be generalizable to all smartphone users. In addition, because of the need for our app to comply with the Australian privacy policy and best practice on users’ confidentiality [62], we were unable to include programming codes within the app that could capture users’ personal profiles such as age, gender, browser, connection speed, etc. Having this information could provide an opportunity to assess different levels of engagement between those who completed the study and those who did not. In addition, we would have been able to assess if app use was moderated by users’ profile. Despite these limitations,
considering the promising results further research with a larger sample and over an extended period of time is necessary.

**Conclusions**

This study provided a comprehensive understanding of participant retention, technology usage, behavioral change process, and engagement with My Care Hub app during a short trial period. Retention was high, although further strategies may be required to further sustain retention when the app is used in long-term trials. The system log indices of FITT of engagement reveal a reasonable level of technology usage during the intervention period. The BIT model employed to measure behavioral change and engagement suggests that My Care Hub could be a behavior change intervention tool to support self-management behaviors in people with type 1 or type 2 diabetes. Information obtained through the use of multicomponent measures of engagement in this study provides rich and useful data regarding the strengths and weaknesses of My Care Hub and areas requiring improvement to foster increased engagement, sustainable long-term use, and effective health behavioral intervention.

**Acknowledgments**

This research was supported by an Australian International Research Training Scholarship and a James Cook University top up scholarship given to the first author. The authors would like to thank Ian Aitkinson and Saira Viqar of the eResearch team at James Cook University for their assistance with the app development. They sincerely appreciate all the study participants and the National Diabetes Service Scheme, Australia, for their invaluable support with the advertisement of the study.

**Authors’ Contributions**

BM, UM, AM, and MA conceived and designed the study. MA and AD collected and analyzed the data. MA prepared the original draft, and BM, UM, AM, and AD reviewed and edited the paper. BM is the project lead. All authors have read and approved the final paper.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Sample screenshots of the My Care Hub app.
[PDF File (Adobe PDF File), 295 KB - mhealth_v8i6e17802_app1.pdf]

Multimedia Appendix 2
Semistructured interview guide.
[PDF File (Adobe PDF File), 413 KB - mhealth_v8i6e17802_app2.pdf]

Multimedia Appendix 3
Consolidated criteria for reporting qualitative research checklist.
[PDF File (Adobe PDF File), 348 KB - mhealth_v8i6e17802_app3.pdf]

**References**

3. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 9;16(4):e104 [FREE Full text] [doi: 10.2196/mir.2968] [Medline: 24718852]


Abbreviations

BG: blood glucose
BGL: blood glucose level
BIT: behavioral intervention technology
DSM: diabetes self-management
e-gift: electronic gift
FITT: frequency, intensity, time, and type
mHealth: mobile health

©Mary D Adu, Usman H Malabu, Aduli EO Malau-Aduli, Aaron Drovandi, Bunmi S Malau-Aduli. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 11.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Using an Interactive App for Symptom Reporting and Management Following Pancreatic Cancer Surgery to Facilitate Person-Centered Care: Descriptive Study

Tina Gustavell¹,², PhD; Kay Sundberg¹, PhD; Ann Langius-Eklöf¹, PhD

¹Karolinska Institutet, Department of Neurobiology, Care Sciences and Society, Division of Nursing, Stockholm, Sweden
²Karolinska University Hospital, Theme Cancer, Department of Upper Abdominal Diseases, Stockholm, Sweden

Corresponding Author:
Tina Gustavell, PhD
Karolinska Institutet
Department of Neurobiology, Care Sciences and Society
Division of Nursing
Alfred Nobels Allé 23
Stockholm, 141 83 Huddinge
Sweden
Phone: 46 707787934
Email: tina.gustavell@ki.se

Abstract

Background: Pancreatic and periampullary cancers are rare but have high mortality rates. The only hope for cure is surgical removal of the tumor. Following pancreatic surgery, the patients have a great deal of responsibility for managing their symptoms. Patients report a lack of sufficient knowledge of self-care and unmet supportive care needs. This necessitates a health care system responsive to these needs and health care professionals who pay close attention to symptoms. Person-centered care is widely encouraged and means a shift from a model in which the patient is the passive object of care to a model involving the patient as an active participant in their own care. To address the challenges in care following pancreatic cancer surgery, an interactive app (Interaktor) was developed in which patients regularly report symptoms and receive support for self-care. The app has been shown to reduce patients' symptom burden and to increase their self-care activity levels following pancreaticoduodenectomy due to cancer.

Objective: The aim of the study was to describe how patients used the Interaktor app following pancreaticoduodenectomy due to cancer and their experience with doing so.

Methods: A total of 115 patients were invited to use Interaktor for 6 months following pancreaticoduodenectomy. Of those, 35 declined, 8 dropped out, and 46 did not meet the inclusion criteria after surgery, leaving 26 patients for inclusion in the analysis. The patients were instructed to report symptoms daily through the app for up to 6 months following surgery. In case of alerting symptoms, they were contacted by their nurse. Data on reported symptoms, alerts, and viewed self-care advice were logged and analyzed with descriptive statistics. Also, the patients were interviewed about their experiences, and the data were analyzed using thematic analysis.

Results: The patients’ median adherence to symptom reporting was 82%. Fatigue and pain were the most reported symptoms. Alerting symptoms were reported by 24 patients, and the most common alert was fever. There were variations in how many times the patients viewed the self-care advice (range 3-181 times). The most commonly viewed advice concerned pancreatic enzyme supplements. Through the interviews, the overarching theme was “Being seen as a person,” with the following 3 sub-themes: “Getting your voice heard,” “Having access to an extended arm of health care,” and “Learning about own health.”

Conclusions: Interaktor proved to be well accepted. It made patients feel reassured at home and offered support for self-care. The app facilitated person-centered care by its multiple features targeting individual supportive care needs and enabled participation in their own care. This supports our recent studies showing that patients using the app had less symptom burden and higher self-care activity levels than patients receiving only standard care.

(JMIR Mhealth Uhealth 2020;8(6):e17855) doi:10.2196/17855
Introduction

Pancreatic and periampullary cancers are rare, with only 1300 individuals (equal proportions of men and women) diagnosed each year in Sweden [1]. The mortality rate is high because surgical resection can only be offered to fewer than 20% of patients [2]. Even after intentionally curative surgery and adjuvant chemotherapy, the prognosis is poor, with a median survival period of 2-4 years depending on whether it is pancreatic or periampullary cancer [3,4]. The most common surgical procedure for these tumors is pancreaticoduodenectomy, which impairs quality of life [5,6]. High demands are put on patients to manage their illness after surgery. It has been concluded that patients who have had a pancreaticoduodenectomy sometimes lack sufficient knowledge of self-care and have unmet supportive care needs, which necessitate a health care system that is responsive to these needs and health care professionals who pay close attention to symptoms [7,8].

Patients must often navigate through a fragmented health care system and adapt to routines customized to the health care organizations and professionals, rather than receiving care designed to focus on the individual patient’s needs, preferences, and values [9]. Person-centered care is today a widely encouraged alternative and means a shift away from a model in which the patient is the passive object of care to a model where arrangements are made involving the patient as an active participant in his or her care [10]. Participation in one’s own care can include mutual communication with health care professionals where patients are listened to and their knowledge is respected, shared knowledge where patients receive explanations of symptoms and procedures and can also tell professionals about their symptoms, and patients knowing how to manage their symptoms and provide self-care [11]. To achieve person-centered care where patients really are active participants, support of a positive attitude to modern innovations is needed. Routine use of patient-reported outcomes in clinical practice can be one way of identifying patients’ current concerns and impact of treatment, enhancing patient-clinician communication, promoting shared decision making, and improving patient satisfaction [12,13]. Medical and public health practices supported by mobile devices have been defined by the World Health Organization as mobile health (mHealth) [14]. It has been reported that patients undergoing cancer treatments who report symptoms to health care professionals through mHealth systems and receive support for symptom management have higher quality of life, less symptom distress [15-17], and improved 2-year survival [18] compared with patients not using such systems.

Given the poor prognosis of pancreatic and periampullary cancer, the distressing symptoms patients experience, and insufficient knowledge of self-care and unmet supportive care needs, challenges arise in supporting patients with cancer following pancreaticoduodenectomy. To address these challenges, an interactive app (Interaktor) for smart devices was developed in which patients regularly report symptoms and receive support through continuous access to self-care advice and their health care professionals. The content in the app was developed by reviewing literature and interviewing patients and health care professionals [19] and has been tested for feasibility [20]. Evaluation of the app’s impact on quality of life has shown higher emotional function and less symptom burden 6 weeks after surgery for patients using the app compared with patients not using the app [21]. Furthermore, patients using the app had higher self-care activity levels 6 months after surgery [21]. Knowledge of the patients’ usage and experience of the app may support the interpretation of these results. Therefore, the aim of the current study was to describe how patients used the Interaktor app following pancreaticoduodenectomy due to cancer and their experience with doing so.

Methods

Design

The current study is part of the evaluation of the Interaktor app adjusted for patients with pancreatic cancer and has a descriptive design. Ethical approval was given by the Regional Ethical Review Board in Stockholm, Sweden (Reg.no: 2011/1780-13/2).

Setting

The study was performed at Karolinska University Hospital, which has the highest volumes of pancreatic surgery in Sweden. Following pancreaticoduodenectomy, at the time of the study, the patients were normally cared for on a surgical ward for 1 to 2 weeks and thereafter at a rehabilitation unit outside the hospital for 1 week. Standard care after discharge was that the patients should contact the clinic’s outpatient unit if they felt the need to. Also, around 5 weeks after surgery, the patients had an appointment with a surgeon at the outpatient unit. After this appointment, patients with a confirmed diagnosis of malignant disease were referred to the oncology clinic to start adjuvant chemotherapy. The chemotherapy had to start within 10 weeks after surgery, and standard treatment was gemcitabine given as an intravenous infusion over 30 minutes, once a week for 3 of every 4 weeks (1 cycle), for 6 cycles.

Sample

During a period of 16 months in 2015-2016, all patients who were scheduled to undergo pancreaticoduodenectomy at the university hospital due to a suspected malignancy in the pancreatic or periampullary region were screened for eligibility. Inclusion criteria were follow-up care planned at the university hospital and able to read and understand Swedish. After the screening process, 115 patients were eligible before surgery. A total of 35 patients declined to participate. After surgery, patients who did not undergo pancreaticoduodenectomy or were too ill were excluded. Upon discharge, 44 patients were introduced to the app. Patients who did not have malignant disease, who died before discharge, who were discharged with advanced home care, or who dropped out were not analyzed, leaving a final sample of 26 patients included in the analysis (Figure 1).

KEYWORDS

pancreatic neoplasms; mHealth; interactive; symptoms; self-care; support; patient-reported outcomes, person-centered care
Characteristics of participants included in the analysis are shown in Table 1.

**Figure 1.** Flowchart of the inclusion process.

**Table 1.** Sociodemographic and clinical characteristics of the study participants (n=26).

<table>
<thead>
<tr>
<th>Sociodemographic and clinical characteristics</th>
<th>Descriptive analyses, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td>67 (8.7)(^a), 67 (51-82)(^b)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>8 (31)</td>
</tr>
<tr>
<td>Male</td>
<td>18 (69)</td>
</tr>
<tr>
<td><strong>Living situation</strong></td>
<td></td>
</tr>
<tr>
<td>Married or living with partner</td>
<td>21 (81)</td>
</tr>
<tr>
<td>Living alone</td>
<td>5 (19)</td>
</tr>
<tr>
<td><strong>Highest education level</strong></td>
<td></td>
</tr>
<tr>
<td>Junior compulsory</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Senior high school</td>
<td>9 (35)</td>
</tr>
<tr>
<td>Postgraduate or university</td>
<td>15 (58)</td>
</tr>
<tr>
<td>Missing data</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Histopathology</strong></td>
<td></td>
</tr>
<tr>
<td>Pancreatic ductal adenocarcinoma</td>
<td>12 (46)</td>
</tr>
<tr>
<td>Periampullary cancer</td>
<td>12 (46)</td>
</tr>
<tr>
<td>Invasive intraductal papillary mucinous neoplasia</td>
<td>2 (8)</td>
</tr>
<tr>
<td><strong>Adjuvant chemotherapy</strong></td>
<td></td>
</tr>
<tr>
<td>Yes, full cycle</td>
<td>17 (65)</td>
</tr>
<tr>
<td>Yes, ceased in advance(^c)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>No</td>
<td>4 (15)</td>
</tr>
</tbody>
</table>

\(^a\)mean (SD).
\(^b\)median (range).
\(^c\)Due to side effects (n=2), recurrent disease (n=2), or death (n=1).

**Interaktor**

The Interaktor app is generic and adjustable depending on the setting and situation. It is designed for both Android and iOS and can be downloaded to any smartphone or tablet and requires a separate log in. The primary features of the Interaktor app are regular assessment of self-reported symptoms, risk assessment models for alerts, continuous access to evidence-based self-care advice and links to relevant websites for more information, and
graphs that allow patients to view their symptom reporting history (Figure 2).

The structure of the symptom assessment was inspired by a standardized symptom questionnaire that assesses a symptom’s occurrence, rated as “yes” or “no,” and a symptom’s frequency and distress level on a 4-point rating scale [22,23]. The pancreas version of Interaktor consists of 12 symptom questions following surgery and 3 additional questions for patients undergoing adjuvant chemotherapy, as defined by patients and health care professionals in our previous studies [19,20]. Patients also have the possibility to write a free-text comment before submitting a report. After completing the symptom assessment, the report is immediately sent to a secure server that is linked to a monitoring web interface where reports and alerts can be viewed. The risk assessment model for alerts is, in this version, programmed differently depending on whether patients undergo chemotherapy. There are two types of alerts: red and yellow. A red alert indicates that the patient is experiencing a severe symptom and should be contacted within 1 hour, and for yellow alerts, contact should be made the same day (Table 2). If an alert is triggered, the patient receives suggestions on self-care advice to read. Further, a text message is automatically sent to a cellphone at the clinic to notify the patient’s nurse to view the alerted symptoms in the web interface.

Figure 2. Screenshots from the Interaktor app adapted for patients following pancreateoduodenectomy showing the primary features: (A) symptom reporting, (B) alerts, (C) self-care advice to read, and (D) graph showing symptom change over the previous week.
Table 2. Risk assessment model for alerts.

<table>
<thead>
<tr>
<th>Symptom alerts</th>
<th>Rated as</th>
<th>Response options</th>
<th>Alert triggered after…</th>
<th>Type of alert</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>After surgery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td>Occurrence</td>
<td>“Yes”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Pain</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Vomiting</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Dizziness</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Vomiting</td>
<td>Frequency</td>
<td>“Sometimes,” “Often,” OR “Almost always”</td>
<td>2 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Dizziness</td>
<td>Frequency</td>
<td>“Sometimes,” “Often,” OR “Almost always”</td>
<td>2 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Loose stool</td>
<td>Frequency</td>
<td>“Often” OR “Almost always”</td>
<td>3 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Constipation</td>
<td>Occurrence</td>
<td>“Yes”</td>
<td>3 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Eating difficulties</td>
<td>Frequency</td>
<td>“Often” OR “Almost always”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Pain</td>
<td>Frequency</td>
<td>“Often” OR “Almost always”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Nausea</td>
<td>Frequency</td>
<td>“Often” OR “Almost always”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Distress</td>
<td>“Rather much” OR “Very much”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Sadness/depression/worry</td>
<td>Distress</td>
<td>“Rather much” OR “Very much”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Problems performing activities at home</td>
<td>Distress</td>
<td>“Rather much” OR “Very much”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td>Problems performing activities outside home</td>
<td>Distress</td>
<td>“Rather much” OR “Very much”</td>
<td>7 consecutive days</td>
<td>Yellow</td>
</tr>
<tr>
<td><strong>During chemotherapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fever</td>
<td>Occurrence</td>
<td>“Yes”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Breathing difficulties</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Nausea</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Vomiting</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Numbness/tingling in hands and/or feet</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Eating difficulties</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Red</td>
</tr>
<tr>
<td>Swelling/pain/redness from SVP/PICC</td>
<td>Occurrence</td>
<td>“Yes”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Loose stool</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Pain</td>
<td>Frequency</td>
<td>“Almost always”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Dizziness</td>
<td>Frequency</td>
<td>“Almost always” OR “Often”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Vomiting</td>
<td>Frequency</td>
<td>“Often”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Nausea</td>
<td>Frequency</td>
<td>“Often”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Breathing difficulties</td>
<td>Frequency</td>
<td>“Often”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Constipation</td>
<td>Distress</td>
<td>“Very much”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
<tr>
<td>Sadness/depression/worry</td>
<td>Distress</td>
<td>“Very much”</td>
<td>One report</td>
<td>Yellow</td>
</tr>
</tbody>
</table>

*Since the patients undergoing chemotherapy have contact with a nurse at least once a week, no alerts were programmed to be triggered after multiple consecutive days.

bSVP: subcutaneous venous port.

cPICC: peripherally inserted central catheter.
Procedure
A researcher helped the patients to download the app to their own smartphone. Patients who did not have access to a smartphone (n=2) were lent one with the app installed. The researcher instructed the patient on the different features; thereafter, the patient practiced submitting a report under the researcher’s supervision. The submitted report was then shown in the graphs and discussed together. The self-care advice, including hyperlinks to websites, was introduced. Furthermore, a written manual for using the app was given to the patients to take home. The patients were instructed to report symptoms daily for at least 4 weeks starting the first day after discharge from the surgical or rehabilitation clinic and up to 6 months after surgery or one week after ceasing adjuvant chemotherapy. After the first 4 weeks of reporting, a researcher called the patients to ask if they wanted to continue using the app. A reminder notification to report was sent through the app every day. The patients were thoroughly informed both orally and in writing that, in case of an alert, they would only be contacted during working hours (8 am to 4 pm on non-weekend days) because the report could only be monitored by a nurse during this time. If an alert was triggered outside of working hours, the patients were called the following weekday.

The patients’ contact nurses were responsible for monitoring alerts. They were employed at the surgical clinic or at the oncology clinic for those patients who underwent adjuvant chemotherapy. The nurses were instructed to call the patients if they received an alert text. One of the researchers could be contacted in the event of any technical problems. Patients who had access to advanced home care with specific home care nurses could not use the app since those nurses were not introduced to the app.

Data Collection
Data concerning the number of submitted reports, reported symptoms, triggered alerts, and viewed self-care advice were logged on a secure server and extracted as an encrypted Excel file.

The patients were interviewed individually after their final report about their experiences with using the app. One patient died within the study period and therefore could not be interviewed. To ensure trustworthiness, the interviews followed a semistructured interview guide with the questions: “What was it like to use the app?” “In which way have you been in contact with health care?” and “In which way have you been able to be involved in your care?” Depending on the extent of the patients’ answers, probing questions like “Can you elaborate or give an example” were used. The interviews lasted for a median time of 31 minutes (range 16-71 minutes) and were audio recorded. To ensure that the patients were comfortable, they were interviewed either in their own home (n=21) or at the hospital (n=4) according to their own choice.

Data Analysis
Logged data from the app were analyzed with descriptive statistics. Adherence to reporting was calculated as the number of days a patient submitted a report divided by the number of days a patient was meant to report and presented as a percentage.

The patients’ interviews were analyzed using thematic analysis, as described by Braun and Clark [24]. First, all interviews were transcribed verbatim and read through several times. Statements regarding the app were systematically coded throughout the entire dataset with an inductive approach. A code could consist of a few words or a whole sentence. Matching codes were then put together and created themes. All data in one theme were then reviewed to see if the theme worked in relation to the codes. This reviewing process was completed by all authors. If a theme did not work, the process of collating codes started from the beginning until all themes worked in relation to the codes and the entire dataset. During the whole process, themes were defined, named, and renamed. Individual quotes were chosen to validate the findings. To establish rigor of the analysis, the 15-point checklist of criteria by Braun and Clark [24] for good thematic analysis was followed [24].

Results
Logged Data
Patients used the app for a median of 190 days (range 35-245 days). The median adherence to reporting daily was 82.2% (range 23.5%-100%). Reasons to stop reporting in advance were their own choice (n=1), follow-up care transferred to unit not included in the study (n=3), or death (n=1).

Reported Symptoms
A total of 6320 symptoms (median 170, range 9-994) were reported, and at the group level, all symptoms were reported but not by each patient (Table 3). The 4-point rating values were all used in the follow-up questions. Levels of frequency and distress of a symptom were mostly concordant except for nausea, vomiting, and dizziness, for which patients reported a higher distress level than frequency and the opposite for numbness in hands or feet (Table 3). Fatigue and pain were the most frequently occurring symptoms and also reported by most patients (Table 3).
Table 3. Occurrences, frequency, and distress of the symptoms as reported in the app by patients (n=26) following pancreaticoduodenectomy due to cancer.

<table>
<thead>
<tr>
<th>Symptoms (number of patients reporting the symptom)</th>
<th>Occurrence (n=6320)</th>
<th>Frequency</th>
<th>Distress</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>Median</td>
<td>Range</td>
</tr>
<tr>
<td>Fatigue (n=24)</td>
<td>1445 (22.86)</td>
<td>35.5</td>
<td>3-198</td>
</tr>
<tr>
<td>Pain (n=23)</td>
<td>863 (13.7)</td>
<td>19</td>
<td>1-169</td>
</tr>
<tr>
<td>Problems performing activities outside home (n=21)</td>
<td>605 (9.6)</td>
<td>21</td>
<td>1-161</td>
</tr>
<tr>
<td>Nausea (n=21)</td>
<td>572 (9.1)</td>
<td>11</td>
<td>1-158</td>
</tr>
<tr>
<td>Eating difficulties (n=22)</td>
<td>535 (8.5)</td>
<td>13.5</td>
<td>2-160</td>
</tr>
<tr>
<td>Loose stool (n=24)</td>
<td>526 (8.3)</td>
<td>6</td>
<td>1-133</td>
</tr>
<tr>
<td>Problems performing activities at home (n=20)</td>
<td>518 (8.2)</td>
<td>12.5</td>
<td>1-127</td>
</tr>
<tr>
<td>Sadness, depression, worry (n=12)</td>
<td>386 (6.1)</td>
<td>14</td>
<td>3-169</td>
</tr>
<tr>
<td>Dizziness (n=15)</td>
<td>267 (4.2)</td>
<td>10</td>
<td>1-91</td>
</tr>
<tr>
<td>Numbness in hands or feet (n=9)</td>
<td>204 (3.3)</td>
<td>2</td>
<td>1-85</td>
</tr>
<tr>
<td>Constipation (n=23)</td>
<td>132 (2.1)</td>
<td>4</td>
<td>1-28</td>
</tr>
<tr>
<td>Fever (n=16)</td>
<td>87 (1)</td>
<td>3</td>
<td>1-18</td>
</tr>
<tr>
<td>Swelling/pain/redness from SVP/PICC (n=9)</td>
<td>69 (1)</td>
<td>3</td>
<td>1-49</td>
</tr>
<tr>
<td>Breathing difficulties (n=7)</td>
<td>61 (1)</td>
<td>3</td>
<td>1-41</td>
</tr>
<tr>
<td>Vomiting (n=14)</td>
<td>50 (0.8)</td>
<td>2.5</td>
<td>1-11</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
bSymptoms only reported during adjuvant chemotherapy.
cSVP: subcutaneous venous port.
dPICC: peripherally inserted central catheter.

Alerts

The total number of alerts was 512 (median 9, range 0-87), and almost all patients (n=24) reported an alert. Of these alerts, 35.5% (182/512) were severe (red). The most common alert was fever, which was also triggered by most patients (Table 4).
Table 4. Distribution of the number of alerts (n=512) reported in the app by patients (n=24) after discharge following pancreaticoduodenectomy due to cancer.

<table>
<thead>
<tr>
<th>Symptom alerts (number of patients generating the alert)</th>
<th>Median (Range)</th>
<th>Red alerts (n=182), n</th>
<th>Yellow alerts (n=330), n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever (n=16)</td>
<td>3 (1-18)</td>
<td>87</td>
<td>N/A^a</td>
</tr>
<tr>
<td>Dizziness (n=13)</td>
<td>5 (0-20)</td>
<td>0</td>
<td>67</td>
</tr>
<tr>
<td>PICC^b (n=9)</td>
<td>3 (1-51)</td>
<td>N/A</td>
<td>72</td>
</tr>
<tr>
<td>Loose stool (n=9)</td>
<td>2 (1-48)</td>
<td>N/A</td>
<td>71</td>
</tr>
<tr>
<td>Nausea (n=8)</td>
<td>2.5 (1-13)</td>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>Pain (n=8)</td>
<td>2.5 (1-5)</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Eating difficulties (n=7)</td>
<td>2 (1-18)</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Constipation (n=7)</td>
<td>1 (1-3)</td>
<td>N/A</td>
<td>10</td>
</tr>
<tr>
<td>Fatigue (n=6)</td>
<td>3 (1-4)</td>
<td>N/A</td>
<td>15</td>
</tr>
<tr>
<td>Problems with activities outside home (n=5)</td>
<td>4 (1-7)</td>
<td>N/A</td>
<td>19</td>
</tr>
<tr>
<td>Vomiting (n=5)</td>
<td>1 (1-1)</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Breathing (n=4)</td>
<td>1.5 (1-6)</td>
<td>N/A</td>
<td>10</td>
</tr>
<tr>
<td>Problems with activities at home (n=3)</td>
<td>1 (1-1)</td>
<td>N/A</td>
<td>3</td>
</tr>
<tr>
<td>Sadness, depression, worry (n=2)</td>
<td>2.5 (2-3)</td>
<td>N/A</td>
<td>5</td>
</tr>
<tr>
<td>Numbness (n=1)</td>
<td>80 (80-80)</td>
<td>80</td>
<td>N/A</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
^bPICC: peripherally inserted central catheter.

Free-Text Comments

The free-text comment section to communicate with health care was used 302 times in total (median 7.5, range 0-90) and used by most patients (n=24). Most comments were a detailed description about a symptom, which was sometimes followed by a wish for counseling or the text “You do not need to call me.” The patients also used the free-text comment section to document values for weight, blood glucose, blood pressure, and temperature or to inform on admission to hospital, going away on holiday, or need for prescriptions.

Self-Care Advice

The patients viewed self-care advice 1231 times in total (median 30.5, range 3-181). The most commonly and least commonly viewed self-care advice is shown in Table 5.

Table 5. The five most and least commonly viewed self-care advice items and number of times viewed by the whole group

<table>
<thead>
<tr>
<th>Self-care advice (number of patients who viewed the advice)</th>
<th>Number of times viewed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Most commonly viewed</strong></td>
<td></td>
</tr>
<tr>
<td>Pancreatic enzyme supplement (n=25)</td>
<td>99</td>
</tr>
<tr>
<td>Dietary advice (n=21)</td>
<td>86</td>
</tr>
<tr>
<td>Pain (n=21)</td>
<td>76</td>
</tr>
<tr>
<td>Fever (n=16)</td>
<td>68</td>
</tr>
<tr>
<td>Weight loss (n=18)</td>
<td>62</td>
</tr>
<tr>
<td><strong>Least commonly viewed</strong></td>
<td></td>
</tr>
<tr>
<td>Sleep disturbance (n=6)</td>
<td>19</td>
</tr>
<tr>
<td>Instable blood sugar (n=12)</td>
<td>19</td>
</tr>
<tr>
<td>Breathing difficulties (n=5)</td>
<td>21</td>
</tr>
<tr>
<td>Hair/skin/mucous membrane (n=5)</td>
<td>27</td>
</tr>
<tr>
<td>Numbness/tingling in hands and feet (n=8)</td>
<td>27</td>
</tr>
</tbody>
</table>
Interviews With Patients

The overarching theme “Being seen as a person” was identified, with the following subthemes: “Getting your voice heard,” “Having access to an extended arm of health care,” and “Learning about own health.” Examples of codes connected to the subthemes are illustrated in Figure 3.

Figure 3. Examples of codes (white ovals) connected to the 3 sub-themes (light gray rectangles) and overarching theme (dark gray rectangle) identified through the thematic analysis of interviews with patients (n=25) using the app following pancreaticoduodenectomy due to cancer.

Being Seen as a Person: Getting Your Voice Heard

The patients overall talked about feeling taken care of at home since someone was keeping an eye on how they were feeling on a daily basis. Patients undergoing chemotherapy noticed that the nurses had viewed their reports and knew how they were feeling when they came to the hospital.

>You just have to send in your report and then you get to talk to someone. /…/ I think you are more involved in care this way since you have your voice heard when you want. [Patient 8]

When the patients had reported an alerting symptom, they expressed how important it was that the nurse called the patient. This conversation with the nurse resulted in individually adjusted advice. The patients felt that they received comfort and help with their problems from the nurses who called after an alert and that they could raise issues other than the reported symptoms.

>Once I could not understand a manual I was reading, my brain wasn’t working, and then I got... afraid. Then I wrote that in the free text and the nurse called and told me about ‘cyto-brain’. It was comforting to get an explanation and to talk about it instead of going around worrying about it alone. [Patient 5]

Being Seen as a Person: Having Access to an Extended Arm of Health Care

The patients described the app as a reassuring link to health care after discharge and that it made them feel like they were still at the hospital even if they were at home. Only having one point of contact and not having to think about who to call was described as a benefit. Patients expressed that they often knew the nurse who contacted them or learned to recognize the voice of nurses they had not met in person. Sometimes the patients did not know or recognize the nurse but the person who called was always well informed. The patients expressed that the app offered a faster and easier way to get in touch with health care than the regular way, which is to call, enter contact information, and either be placed on hold or called back later. The need for fast and easy contact with health care was most prominent during the first weeks when a lot of symptoms were present but also later if new symptoms arose due to chemotherapy.

>Someone is checking up on you, so that you are not starting to feel too bad. And that is great. It’s something in-between being at home and lying in hospital, but at home you are free. [Patient 12]

Using the app was a sort of follow-up that was otherwise lacking, and more contact with health care was initiated since they would not themselves have called as often if the decision was theirs. Sometimes the patients were not contacted after an
The app decides when you will be contacted and that feels a bit weird, because sometimes it’s okay, but sometimes it’s not okay, and then the nurse and I agreed that when I reported symptoms and didn’t want, or need, to be contacted, I wrote that in the free text. [Patient 16]

Other patients found it reassuring that someone else was responsible for making the decision if contact was needed and knowing that if someone did not call, everything was satisfactory.

The app made me not have to judge myself what is cause for concern. Instead I could leave that to someone else. Not having to think about if it was something I needed to react to, but instead just hand myself over. [Patient 17]

**Being Seen as a Person: Learning About Own Health**
The patients could identify important symptoms and reflect on how they felt since they were asked to rate symptoms daily.

To think about how you feel every day is a perspective that I think is especially beneficial, because it is very easy to think that you are completely well and then you push yourself too much. [Patient 4]

Some patients thought it was helpful to analyze their symptom change over time when symptoms had been unstable.

I was curious to see if my symptoms, like lack of appetite and tiredness, were connected to the treatment. And it seems that the day after treatment, and the following two or three days, then the tiredness is at its worst, whereas changes in appetite are much slower. [Patient 5]

Having access to self-care advice provided new knowledge on symptoms and how to manage them and gave explanations as to why they were feeling as they were, and misconceptions could be dealt with.

I think reading the advice has been valuable to be able to justify, why it’s true, why I feel like I do, or if there is something I need to think about. [Patient 18]

The patients expressed that having easy access to the advice was important since information is easy to forget, they might not have been given enough information before discharge, or they were not able to absorb information at that time.

I understand that the staff don’t have time to explain everything, or that you are not in the right frame of mind to understand everything they tell you. It was good to have the app directly after being discharged following the surgery and at that time I used that self-care feature a lot /.../ when you Google you can end up on strange sites that don’t reflect your situation so this was more straightforward and concise and contains 100% facts. [Patient 5]

**Discussion**

**Principal Findings**

This study shows that using an interactive app for symptom reporting and management is accepted by patients who have undergone pancreaticoduodenectomy due to cancer and enables person-centered care after discharge. The findings confirm the intent of Interaktor to offer a support system that provides several features that address individual supportive care needs. Our previous results have shown that patients who used the app experienced higher emotional function, less symptom burden, and higher self-care activity levels after surgery compared to patients not using the app [21], which is supported by the results in the present study. There was large variation in how patients used the app and interacted with the nurses, for example, how they wrote free-text comments and viewed self-care advice. Irrespective of how the patients used the app’s features, their experiences with using it were similar. The patients described how the app gave them reassurance in being monitored and having contact with health care, as well as receiving support for self-care.

**Limitations**

Although 115 patients were approached before surgery, data from only 26 patients could be analyzed. Many patients were not eligible upon discharge due to the severity of the disease or treatment, showing the complexity with including this patient group in clinical trials. The initial consent rate was high, specifically 69.6% (80 consented of 115 approached), a rate comparable to a feasibility study of a similar intervention [25]. The consent rate might have been even higher if patients were approached upon discharge when they are more focused on their need for supportive care at home. Patients who declined to participate in the study may have been less interested or experienced obstacles in using a smartphone app compared to those patients who consented. Interest and ability to use mHealth are likely to constantly grow as smartphone access is increasing every year. For instance, recent mapping shows that 90% of the Swedish population have access to a smartphone [26]. Some patients brought up that they had forgotten about some of the features of the app that must be considered when interpreting the results. In future studies of Interaktor and other mHealth tools, it is advisable to make time for a number of training opportunities. In this study, monitoring of and response to alerts could only be made during working hours on weekdays due to the organizational structure at the participating clinics. In future studies, and especially if the app should be implemented in standard care, monitoring of and response to alerts should be
made at all hours of the day and not just restricted to certain hours.

Comparison to Prior Work
The patients had a median adherence of 82.2% for reporting symptoms, which can be considered high, especially since the reporting period was 6 months. Some patients even reached 100% adherence, meaning that they reported symptoms every day for 6 months. This is a major strength of this study and shows the participating patients’ interest and need to use the app. The high median adherence rate has been shown in patients with prostate cancer using Interaktor during radiotherapy treatment [27]. All assessed symptoms were reported in the app, and the patients perceived that the questions covered all experienced symptoms and that specifications could be made in the free text if needed. Interestingly, the patients’ responses on the 4-point rating scale of a symptom’s frequency and distress level were concordant for most symptoms. The coherent responses indicate that it is enough to ask for symptom occurrence, rated by “yes” or “no,” and then either frequency or distress, an approach previously evaluated to be sufficient [28]. Not only do the findings provide knowledge about which symptoms patients normally experience following pancreaticoduodenectomy, they also show that there is large individual spread between symptom experiences. Likewise, there was large spread in how many alerts the patients triggered. Of the reported symptoms, 8% triggered an alert. Even so, none of the patients in the present study expressed that alerts were triggered too seldom. On the contrary, a few patients felt that alerts had been triggered when they felt no need to be contacted. They had then learnt to adjust their responses so that an alert would not be triggered or used the free text to communicate if they did not want to be contacted, a strategy also described by patients with prostate cancer [27]. The possibility to write a free-text message was highly used and appreciated, not only to communicate whether contact was needed but also to raise other needs. Based on the results, the risk assessment model seems adequate for patients with pancreatic cancer with the added possibility to write a free-text message. At a group level, all self-care advice included in the app was viewed, although there was large variety in how often patients viewed the advice. The findings show a pattern where the most occurring symptoms are linked to the most viewed self-care advice. This shows the importance of having advice connected to experienced symptoms and that the app targets individual needs.

Person-centered care is defined as shifting the focus from the disease to the person with the illness — a person with individual needs and preferences — and by doing so, the person can be engaged as an active partner in his or her own care and treatment [9]. The results show several ways in which the app facilitates person-centered care by targeting individual needs, namely, by viewing self-care advice as often as needed and connected to experienced symptoms, communicating to the nurse through free-text messages, analyzing one’s own symptoms, getting individual advice following an alert and call by the nurse, and experiencing an easier way to contact health care. By targeting these needs, the contact and care after discharge can be tailored to the patient’s needs and preferences and not according to a standardized disease-specific schedule.

The findings that the patients got support for symptom management and felt reassured in being monitored and having an easy way to stay in contact with health care are consistent with experiences from patients with other types of cancer [27,29,30]. Moreover, being monitored and contacted after submitting a report has been experienced as participating in one’s own care by patients with colorectal cancer who used a cellphone-based system to report side effects during chemotherapy treatment to health care providers [31] and by patients with prostate cancer using Interaktor to report side effects during radiotherapy [32]. In this study, patients’ participation was also evident when patients made agreements with their nurse as to when they needed to be contacted or adjusted their responses when they did not want to be contacted. Also, patients created relationships with the nurses and shared knowledge and information in connection with an alert and increased their own knowledge by viewing self-care advice, aspects determined to be vital for patient participation in previous studies [33]. It has been stated that illness and poor health could hinder patient participation [33]. However, in the interviews, patients described that the need to use the app was most relevant during times when they felt most unwell. As such, using an app like Interaktor can support patients with poor health to enhance their wellbeing and participate in their own care.

Most discrepancies in opinions about the app concerned the text message that was automatically sent to a nurse if an alert was triggered. Some patients wanted to decide for themselves whether to be contacted while others thought that it was reassuring to know that the decision was somebody else’s. Considering previous results showing that cancer survivors feel unable to judge the seriousness of their symptoms [34], it does not seem to be wise to lay the full responsibility for contact with health care on the unwell patient. However, in further adjustments of the app, these opinions need to be addressed, for instance by offering patients an easy way to communicate whether they wish to be contacted and the reason why.

Patients in this study did not feel there were any negative aspects in answering questions about symptoms. On the contrary, it was found helpful to identify important symptoms and reflect on how they felt in a rational and conscious manner. Similar positive statements have been made by patients with prostate cancer using Interaktor [27]. However, these experiences are in contrast to patients using another self-reporting cellphone-based system where answering questions about side effects of treatment sometimes made patients aware of their side effects in a negative way, causing upsetting emotions [31]. The discrepancies could be due to patients in the latter study not being able to view self-care advice in connection with their reported symptoms.

Conclusion
The Interaktor app proved to be well accepted by patients following pancreaticoduodenectomy due to cancer. It made patients feel reassured at home and offered support for self-care. Also, the app facilitated person-centered care through its multiple features targeting individual supportive care needs and enabled participation in own care. This supports our recent studies showing that patients using the app had less symptom...
burden and higher self-care activity levels than those only getting standard care. This study shows that there are good reasons to implement mHealth support systems for patients with pancreatic cancer.

Acknowledgments

The authors thank the patients who participated in this study and the nurses who monitored the patients and responded to alerts. We also thank Birgitta Holmgren, research nurse, for help in identifying patients who could be asked to participate in the study. Furthermore, we thank Health Navigator for collaboration throughout this study and the Swedish Cancer Foundation, Karolinska Institutet, and Karolinska University Hospital for funding.

Conflicts of Interest

None declared.

References


Enhancement of Neurocognitive Assessments Using Smartphone Capabilities: Systematic Review

John Michael Templeton¹, MSc; Christian Poellabauer¹, PhD; Sandra Schneider², PhD

¹Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, United States
²Department of Communicative Sciences and Disorders, Saint Mary’s College, Notre Dame, IN, United States

Corresponding Author:
John Michael Templeton, MSc
Department of Computer Science and Engineering
University of Notre Dame
384 Fitzpatrick Hall
Notre Dame, IN, 46556-5637
United States
Phone: 1 4802968562
Email: jtemplet@nd.edu

Abstract

Background: Comprehensive exams such as the Dean-Woodcock Neuropsychological Assessment System, the Global Deterioration Scale, and the Boston Diagnostic Aphasia Examination are the gold standard for doctors and clinicians in the preliminary assessment and monitoring of neurocognitive function in conditions such as neurodegenerative diseases and acquired brain injuries (ABIs). In recent years, there has been an increased focus on implementing these exams on mobile devices to benefit from their configurable built-in sensors, in addition to scoring, interpretation, and storage capabilities. As smartphones become more accepted in health care among both users and clinicians, the ability to use device information (eg, device position, screen interactions, and app usage) for subject monitoring also increases. Sensor-based assessments (eg, functional gait using a mobile device’s accelerometer and/or gyroscope or collection of speech samples using recordings from the device’s microphone) include the potential for enhanced information for diagnoses of neurological conditions; mapping the development of these conditions over time; and monitoring efficient, evidence-based rehabilitation programs.

Objective: This paper provides an overview of neurocognitive conditions and relevant functions of interest, analysis of recent results using smartphone and/or tablet built-in sensor information for the assessment of these different neurocognitive conditions, and how human-device interactions and the assessment and monitoring of these neurocognitive functions can be enhanced for both the patient and health care provider.

Methods: This survey presents a review of current mobile technological capabilities to enhance the assessment of various neurocognitive conditions, including both neurodegenerative diseases and ABIs. It explores how device features can be configured for assessments as well as the enhanced capability and data monitoring that will arise due to the addition of these features. It also recognizes the challenges that will be apparent with the transfer of these current assessments to mobile devices.

Results: Built-in sensor information on mobile devices is found to provide information that can enhance neurocognitive assessment and monitoring across all functional categories. Configurations of positional sensors (eg, accelerometer, gyroscope, and GPS), media sensors (eg, microphone and camera), inherent sensors (eg, device timer), and participatory user-device interactions (eg, screen interactions, metadata input, app usage, and device lock and unlock) are all helpful for assessing these functions for the purposes of training, monitoring, diagnosis, or rehabilitation.

Conclusions: This survey discusses some of the many opportunities and challenges of implementing configured built-in sensors on mobile devices to enhance assessments and monitoring of neurocognitive functions as well as disease progression across neurodegenerative and acquired neurological conditions.

(JMIR Mhealth Uhealth 2020;8(6):e15517) doi:10.2196/15517

KEYWORDS
mobile phone; mobile health; neurocognitive tests; neurodegenerative disease; neurocognitive disorders
**Introduction**

In recent history, a crossover between the fields of personal health care and mobile technology has been developed [1]. According to a 2015 US national survey on health-related apps among mobile phone owners [2], over 58% of participants had downloaded a health-related mobile app to focus on health, fitness, or medical care. This suggests that people with mobile devices not only care about their health but are also willing to use their mobile technology to help track and monitor their health in a multitude of ways. Similarly, a study [3] depicts both the American Physical Therapy Association and American Occupational Association advocating the integration of mobile health apps and systems into clinical practice, suggesting that mobile technology is also gaining clinical traction and relevance. As mobile devices become more commonplace in the health space, the formation of new and more robust health apps should be a focus.

This paper aims to provide a systematic analysis by (1) providing background on neurocognitive conditions, functional areas, and their subcategories; (2) understanding mobile technology for the purpose of updating and enhancing traditional assessment tools; (3) discussing challenges and opportunities; and (4) providing a description of a comprehensive mobile assessment tool that both individuals and clinicians can use to monitor wellness and/or decline with respect to neurocognitive function.

In this paper, we follow the Merriam-Webster’s medical definition of neurocognition: “of, relating to, or involving cognitive functioning and associated structures and processes of the central nervous system (the part of the nervous system which in vertebrates consists of the brain and spinal cord, to which sensory impulses are transmitted and from which motor impulses pass out, and which supervises and coordinates the activity of the entire nervous system).” Note that many neurological diseases and conditions yield subsequent cognitive impairments, and functional tests monitor both neurological and cognitive processes. Neurocognitive allows for the description of both.

Neurocognitive assessments are relevant and necessary for evaluating and monitoring neurological diseases across the categories of neurodegenerative [4], neurodevelopmental [5], neuropsychological [6], and traumatic brain injuries (TBIs) [7] or acquired brain injuries (ABIs) [8]. Neurodegenerative conditions present with progressive degeneration of neurons and neural structures. Examples include Parkinson disease, dementia, and amyotrophic lateral sclerosis [4]. Neurodevelopmental conditions (eg, autism spectrum disorders, Down syndrome, and attention deficit hyperactivity disorder) come from complications in the development of the brain [5]. TBIs, such as concussions and chronic traumatic encephalopathy, can occur in a variety of ways [7]. ABIs include stroke and meningitis [8]. Neuropsychological conditions present with behavioral and/or emotional changes, which could be the result of brain damage or a traumatic experience (eg, depression, anxiety, and post-traumatic stress disorder) [6]. Conditions could yield similar presentations to others; however, each category has unique onset conditions. Neurological diseases and conditions and their presentations that may occur are shown in Figure 1. Note that not every condition will present with all the features of that specific disorder. Combinations of symptoms may manifest depending on the individual; their age; socioeconomic background; as well as the stage, severity, and progression of the disease. Regardless of the onset conditions, understanding the taxonomies of the variety of neurocognitive conditions is vital for doctors and clinicians to formulate and administer assessments for correct diagnoses, monitoring, and rehabilitation.

**Methods**

**Assessment of Neurocognitive Functions**

Neurocognitive functions of interest include motor, memory, speech, language, executive function, sensory, behavioral and psychological, sleep, and autonomic functions (Figure 1). Each of these functions correspond generally to various regions of the brain, as can be seen in Figure 2 along with their respective subfunctions. However, these brain regions are multifunctional in nature; thus, functions of interest are closely integrated with each other, the nature of which is not currently completely understood [9].

There are currently formal clinical tests that can be used either for screening or assessing some of these functions of interest depicted in Table 1. Screening assessments such as the Mini-Mental Status Evaluation and Montreal Cognitive Assessment provide a quick general assessment of an individual with suspected neurocognitive impairment and identify areas needing further comprehensive evaluation. These assessments focus on a range of neurocognitive functions [10,11]. More comprehensive assessments such as the Boston Diagnostic Aphasia Examination, Dean-Woodcock Neuropsychological Assessment System, and Neurobehavioral Functioning Inventory aim to assess additional components or assess to a deeper extent [12,13]. However, none of these assessments include all functional areas of interest. A further breakdown of clinical...
screenings and assessments at the test level is shown in Table 2. In addition, Table 3 is a brief collection of studies and reviews across categorical neurocognitive conditions, relevant neurocognitive functions, and functional tests. Traditional testing methods for each neurocognitive function can be understood using Tables 2 and 3 and Figure 3.

As mobile devices are becoming more commonplace in neurocognitive assessments, it is necessary to review device sensors and interactions that are useful for the collection of relevant and objective data. Although some higher-end mobile devices may have additional on-device capabilities and/or sensors, currently all smartphone devices have the minimum set of capabilities listed in Table 4. Utilizing these device-based sensors and/or interactions in the formation and configuration of functional tasks enhances the usefulness and quality of the data. With the increased opportunity for user participation on their own devices and the ability of the clinician to collect and analyze enhanced objective datasets, this becomes a robust modality for the administration of these neurocognitive assessments.

Figure 2. Neurocognitive breakdown into subcategories for a detailed and comprehensive assessment.

Table 1. Current tests assessing functions of interest.

<table>
<thead>
<tr>
<th>Functions</th>
<th>Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MMSE&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Motor</td>
<td>X&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Memory</td>
<td>X</td>
</tr>
<tr>
<td>Speech</td>
<td>X</td>
</tr>
<tr>
<td>Language</td>
<td>X</td>
</tr>
<tr>
<td>Executive</td>
<td>X</td>
</tr>
<tr>
<td>Sensory</td>
<td>_&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>Behavioral</td>
<td>—</td>
</tr>
<tr>
<td>Sleep</td>
<td>—</td>
</tr>
<tr>
<td>Autonomic</td>
<td>—</td>
</tr>
</tbody>
</table>

<sup>a</sup>MMSE: Mini-Mental Status Evaluation.
<sup>b</sup>MoCA: Montreal Cognitive Assessment.
<sup>c</sup>BDAE: Boston Diagnostic Aphasia Examination.
<sup>d</sup>NFI: Neurobehavioral Functioning Inventory.
<sup>e</sup>DWNAS: Dean-Woodcock Neuropsychological Assessment System.
<sup>f</sup>X denotes there is a cross-section between a clinical test and an assessment of the corresponding function.
<sup>g</sup>There is no cross-section between a clinical test and an assessment of the corresponding function.
### Table 2. Test types and their functionalities.

<table>
<thead>
<tr>
<th>Test</th>
<th>Basic functionality of test</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Recall</td>
<td>Recall prompted words from memory</td>
<td>[14]</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>Quantify time to recognize change in state</td>
<td>[15]</td>
</tr>
<tr>
<td>Static Balance</td>
<td>Assess stability and sway in static positions</td>
<td>[16]</td>
</tr>
<tr>
<td>Sit to Stand</td>
<td>Gross motor analysis to and from static positions</td>
<td>[17]</td>
</tr>
<tr>
<td>Functional Gait</td>
<td>Gross motor analysis of gait patterns</td>
<td>[18]</td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>Perform motor sequences across body location</td>
<td>[19]</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>Assess ability to inhibit cognitive interference</td>
<td>[20]</td>
</tr>
<tr>
<td>Wechsler Memory Scale</td>
<td>Recreate visual patterns or heard sequences</td>
<td>[21]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>Sort cards based on changes in stimulus conditions</td>
<td>[22]</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>Connecting objects based on a given set of parameters</td>
<td>[23]</td>
</tr>
<tr>
<td>Bender-Gestalt Test</td>
<td>Reproduce images or patterns from various prompts</td>
<td>[24]</td>
</tr>
<tr>
<td>Spatial Orientation</td>
<td>Orientation or manipulation of objects based on direction</td>
<td>[25]</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>Name common objects following visual cues</td>
<td>[26]</td>
</tr>
<tr>
<td>Syllable Repetition</td>
<td>Repeat various syllables or sequences</td>
<td>[27]</td>
</tr>
</tbody>
</table>

### Table 3. Collection of relevant studies for traditional testing.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Category</th>
<th>Condition</th>
<th>Participants and reviews</th>
<th>Function(s)</th>
<th>Test(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbosa et al [28]</td>
<td>Degenerative</td>
<td>Parkinson’s disease</td>
<td>40 Parkinson’s disease and 45 control</td>
<td>Executive function and speech</td>
<td>Trail Making Test and verbal and semantic fluency</td>
</tr>
<tr>
<td>Levenson et al [29]</td>
<td>Degenerative</td>
<td>Alzheimer’s disease and dementia</td>
<td>Review of papers</td>
<td>Emotion and behavioral and psychological</td>
<td>Structured interviews, rating scales, questionnaires, and behavioral observations</td>
</tr>
<tr>
<td>Rocchi et al [30]</td>
<td>Degenerative</td>
<td>Parkinson’s disease</td>
<td>27 Parkinson’s disease</td>
<td>Autonomic function and motor</td>
<td>Head - up tilt test, Valsalva maneuver, deep breathing, and handgrip test</td>
</tr>
<tr>
<td>Czuba et al [13]</td>
<td>TBIa</td>
<td>TBI</td>
<td>108 Post TBI</td>
<td>Memory, motor, speech, sensory, executive function, and behavioral and psychological</td>
<td>Self-reporting Neurobehavioral Functioning Inventory tool (depression, somatic, memory, attention, communicate, aggression, and motor)</td>
</tr>
<tr>
<td>Whyatt et al [31]</td>
<td>Developmental</td>
<td>Autism</td>
<td>18 Autism, 19 control in group 1, and 22 control in group 2</td>
<td>Motor</td>
<td>Catching a ball (reflex, gross motor, and fine motor) and static balance</td>
</tr>
<tr>
<td>O’Hearn et al [32]</td>
<td>Developmental</td>
<td>Autism</td>
<td>Review of papers</td>
<td>Executive function and memory</td>
<td>Spatial orientation tasks, working memory, response, and inhibition</td>
</tr>
<tr>
<td>Langhorne et al [33]</td>
<td>ABIb</td>
<td>Stroke</td>
<td>Review of papers (Average of 70 subjects per trial)</td>
<td>Motor</td>
<td>Sit to Stand, standing balance, gait, gross motor (arm), and fine motor (hand)</td>
</tr>
<tr>
<td>Brady et al [34]</td>
<td>ABI</td>
<td>Stroke</td>
<td>Review of papers (3002 subjects)</td>
<td>Speech and language</td>
<td>Speech and language and therapies</td>
</tr>
<tr>
<td>Johnsen et al [35]</td>
<td>Psychological</td>
<td>PTSDc</td>
<td>Review of 28 studies</td>
<td>Speech and memory</td>
<td>Wechsler Memory Scale Auditory Verbal Learning Test and California Verbal Memory Test</td>
</tr>
<tr>
<td>Goldstein et al [36]</td>
<td>Psychological</td>
<td>PTSD and depression</td>
<td>Review of studies</td>
<td>Sleep and emotion</td>
<td>Mood Scales, diary documentation, and questionnaires</td>
</tr>
</tbody>
</table>

aTBI: traumatic brain injury.  
bABI: acquired brain injury.  
cPTSD: post-traumatic stress disorder.
Table 4. References to previous publications regarding mobile device sensors and/or capabilities to monitor neurocognitive functions of interest.

<table>
<thead>
<tr>
<th>Device Capabilities</th>
<th>Functions</th>
<th>Motor</th>
<th>Memory</th>
<th>Speech</th>
<th>Language</th>
<th>Executive</th>
<th>Sensory</th>
<th>Behavioral</th>
<th>Sleep</th>
<th>Autonomic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accelerometer</td>
<td>Yang et al [37] and Mathie et al [38]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Gyroscope</td>
<td>Yang et al [37]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GPS</td>
<td>Cavallo et al [41]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Microphone</td>
<td>—</td>
<td>Vacher et al [42]</td>
<td>Rosenblum et al [43] and Vacher et al [42]</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Metadata input</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>Fries et al [52]</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Lock, unlock and app usage</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

*There is no cross-section between device capabilities and corresponding functions.*
Results

Motor

Background and Subcategories

Completing motor tasks are often difficult for individuals with neurodegenerative conditions [56,57], neurodevelopmental conditions [31], and TBI and/or ABIs [33,58]. Motor functions can be subcategorized into fine motor, reflexes, balance, and gross motor. Traditional functional tests for the assessment of motor function are shown in Tables 3 and 5. Fine motor function testing involves the movement of the small muscle groups in one’s hands, fingers, and wrists. Methods for testing these movements include both written tests (eg, using pen and paper for trail making or writing) and object manipulation and/or interaction (eg, orienting an object in space or interacting with cards). Reflex testing requires a quick reaction motor response to an outside stimulus, which can be tactile, visual, and/or aural stimulation (eg, catching an object). Balance testing examines the user’s ability to distribute weight evenly, enabling them to remain steady. This can be examined statically (eg, standing on one leg) or dynamically (eg, going from a seated to a standing position). Gross motor function testing involves the movement of the large muscle groups for functional mobility (eg, gait). Note that although some of the tests listed in Table 5 (eg, Spatial Orientation Tests, the Trail Making Test, and the Wisconsin Card Sorting Test) are not specifically motor tests, the manner in which responses are collected allows the isolation of motor performance metrics.

Table 5. Motor functional tests and assessment methods.

<table>
<thead>
<tr>
<th>Motor subcategories and functional tests</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine motor</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>[23]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>[22]</td>
</tr>
<tr>
<td>Reflex</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Reaction Time Tests</td>
<td>[15]</td>
</tr>
<tr>
<td>Balance</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Static Balance</td>
<td>[16]</td>
</tr>
<tr>
<td>Sit to Stand</td>
<td>[17]</td>
</tr>
<tr>
<td>Gross motor</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Functional Gait Assessment</td>
<td>[18]</td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
</tbody>
</table>

Mobile Assessments

Mobile testing and analysis of motor functions use a variety of human-device interactions [50] and positional (eg, accelerometer, gyroscope, and GPS) [37,38,41], media (eg, camera) [46], and inherent (eg, device timer) [49] sensors, as shown in Table 4. Each motor subfunction calls for a subset of device capabilities to gain additional concrete metrics aiding in monitoring, diagnosis, and rehabilitation. Human-device interactions utilize sensors to monitor the positional state of a user’s finger during a tracing task on the screen, either via electrical current or reflection of waves. The output of this can be expressed as coordinates in 2-dimensional space and/or force measurements [50]. An example includes a geometric object (eg, circle or square) being displayed on the screen with the user’s intention to trace the shape. Relative coordinates of the trace path compared with the coordinates of the actual shape provide specific objective metrics (eg, the number of times the outline was crossed or the average distance of the trace from the outline). Positional sensors (accelerometers and gyroscopes) are used to capture device motion (eg, when a user moves the device, linear and rotational motion can be assessed) [37,38]. These sensors can be helpful in enhancing object manipulation testing (eg, having the subject manipulate the mobile device itself), balance (eg, monitoring for the lack of linear and rotational motion), and gross motor function (eg, placing the device on the subject’s center of mass for gait assessment). Gross motor function can also employ the device’s GPS capabilities for additional positional information [41]. A device camera can aid in the assessment of motor function both qualitatively and quantitatively. Video analysis techniques such as slow motion or stop-action viewing can be helpful for the qualitative analysis of movement. Quantitative motion analysis of exercise activities can be performed with a detailed analysis of video recordings to analyze the subject’s movement patterns [59]. The inherent device timer allows for temporal metrics to be collected in conjunction with each of the previously
mentioned metrics (eg, maximum speed, average speed, and acceleration) [49]. Finally, the sampling rate of the device’s sensors can be configured to collect additional data points as needed for an objective fine-grained analysis of the motor function.

A few mobile device apps currently being used for motor assessment include gait feedback and activity recognition. In gait rehabilitation and training, mobile device sensors were used to collect metrics for the analysis of gait patterns to establish corrective adjustments [60]. This study [60] implemented sensory feedback based on gait metrics and monitored how that feedback was interpreted for the change of subsequent steps. Similarly, mobile device sensors can aid in the classification of activities [61]. Requiring a subject to wear a smartphone on their waist to collect accelerometer and gyroscope data while various activities are performed allows for activity classification metrics to be collected.

**Memory**

**Background and Subcategories**

Memory is another sector of neurocognitive assessment prominent in neurodegenerative conditions [62,63], neurodevelopmental conditions [64], TBIs or ABIs [65], and neuropsychological conditions [35]. Memory analysis can be broken down into short-term or working memory, long-term memory, and skill memory. Natural fluctuations in memory based on stress and/or fatigue are normal; however, continual trends over time showing overall decline are important for the diagnosis of diseases. Traditional functional tests regarding memory functions are presented in Tables 3 and 6. Short-term or working memory is the ability to maintain a small amount of basic information for a short period. User comprehension of a simple set of instructions, remembering visual patterns or auditory cues, are all ways in which short-term or working memory can be assessed. Long-term memory is the ability to maintain information over a long period. This information can be provided to the user via verbal, visual, or written modes. Assessment of this information could include recalling an event from a user’s past (eg, episodic memory) or could require the user to memorize information for a later assessment. Skill memory requires the individual to carry out normal functions and/or interactions without requiring much thought (eg, riding a bike or driving a car).

**Table 6. Memory functional tests and assessment methods.**

<table>
<thead>
<tr>
<th>Memory subcategories and functional tests</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short-term or working</td>
<td></td>
</tr>
<tr>
<td>Bender-Gestalt Test</td>
<td>[24]</td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
<tr>
<td>Wechsler Memory Scale</td>
<td>[21]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>[22]</td>
</tr>
<tr>
<td>Long-term</td>
<td></td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
<tr>
<td>Wechsler Memory Scale</td>
<td>[21]</td>
</tr>
<tr>
<td>Word Recall Test</td>
<td>[14]</td>
</tr>
<tr>
<td>Skill</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
</tbody>
</table>

**Mobile Assessments**

Mobile testing and analysis of memory function make use of participatory device interactions. Human-device interactions [50], device microphones [42], and inherent [49] sensors are used for the assessment of subfunctions of short-term and long-term memory, as the user must engage with the device providing information they are intended to remember (Table 4). Human-device interactions for memory function could be used to help depict the number of times a user interacts with the screen to gather necessary information to complete a task for short- or long-term memory assessment. A mobile device enhancement of a spatial orientation game, as seen in Figure 3, comprised screen interactions notating visual cues to depict the user flipping over a card to match an original pattern. Media sensors with speech recognition capabilities can also be used for both short-term and long-term memory assessments, such as word or event recall (eg, using speech recognition for certain keywords). Subject interactions for the assessment of skill memory using Apraxia tests [19] may use positional [37,38] or media [46] sensors in a similar manner to motor function (Table 4). Having the subject wear the device while completing a physical skill task (eg, riding a bike) would yield positional metrics for balance and gross motor function to show their overall capability in the task. Skill memory in the form of explaining a procedure (eg, how to make a peanut butter and jelly sandwich) would require the device microphone or human-device interactions. The device timer is highly important for the assessment of memory function, helping to depict the length of time the user takes to express retained information.

A current mobile device app for monitoring memory function involves a memory game for rehabilitation and training, following an ABI [66]. This game hinges on a classic card matching concept in which the user must flip the cards over in
pairs attempting to match all cards with their mates. Metrics are then analyzed with respect to user interactions in the app to track user memory.

**Speech**

**Background and Subcategories**

Speech has become increasingly useful for the purpose of disease diagnostics. Variations in speech could be used as indicators of neurocognitive impairments across the categories of neurodegenerative [28], neurodevelopmental [67], and TBI and/or ABIs [68]. References to traditional speech testing methods are depicted in Tables 3 and 7. Speech analysis is typically broken down into frequency measures and their variations, stress, and repeatability. The fundamental frequencies and variations are acoustic characteristics of speech. Stress in speech is the degree of emphasis given a sound or syllable that can help distinguish the meanings of words or phrases. Repeatability in speech is the ability to replicate syllabic sequences for quickness and accuracy metrics.

<table>
<thead>
<tr>
<th>Speech subcategories and functional tests</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Bosting Naming Test</td>
<td>[26]</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>[20]</td>
</tr>
<tr>
<td>Stress</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Bosting Naming Test</td>
<td>[26]</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>[20]</td>
</tr>
<tr>
<td>Repeatability</td>
<td></td>
</tr>
<tr>
<td>Apraxia Tests</td>
<td>[19]</td>
</tr>
<tr>
<td>Syllable Repetition Test</td>
<td>[27]</td>
</tr>
</tbody>
</table>

**Mobile Assessments**

References for speech analysis on a mobile device using the device’s built-in microphone [42,43] and/or camera [43] to gather sound recordings are shown in Table 4. When collected, these recordings can be used to analyze additional and deeper metrics of the speech sample. This can be expressed with measurable hertz values (eg, fundamental frequencies and their variations). The device’s speech recognition capabilities can be used for the assessment of syllable repetition tests or for evaluating what the user is saying. Each of these modes utilizes the device’s timer for the corresponding temporal metrics of the speech sample [49]. In the enhancement of a syllable repetition test (eg, repeating the sequence of Pa-Ta-Ka after a single deep breath), sensors detect metrics of accuracy (eg, number of correct sequences said by the user), frequency (eg, starting and ending frequencies), and time (eg, how long the user sustained the speech pattern).

Current mobile apps for speech function include the diagnosis, monitoring, and treatment of individuals with speech disorders [68,69]. Many of these speech apps are best suited for user difficulties in phonological representation, articulation, and phonotactics.

**Language**

**Background and Subcategories**

Language is important for the assessment of phonology, morphology, semantics, syntax, and pragmatics. Phonology is the study of phonemes (eg, most basic speech sounds) of an individual language. Morphology is the study of words and other meaningful units of language. Semantics is the study of sentence meaning. Syntax is the study of sentences and phrases and the rules of grammar that they obey. Finally, pragmatics is the study of sentence meanings in context. These fundamental components of language are instrumental in assessing all neurocognitive classifications [67,70-72]. Language assessments look at how the user applies and arranges words, as well as connotation and context, into conversation, presupposition, implication, and overall systematic organization of these words [73]. A reference for the traditional assessment of language is given in Table 3.

**Mobile Assessments**

The references in Table 4 show how mobile assessments can evaluate language using a device’s built-in microphones [42] to gather sound recordings and speech recognition capabilities. These recordings can be analyzed for their linguistic style at each level of the language spectrum (eg, sentences in context or recall based on generated cues [73]). User-device interactions (eg, screen swipes or clicks) can be used for word ordering or comprehension tasks [50]. By enhancing a speech-language task (eg, picture description) on a mobile device, speech recognition can be used to assess word ordering, tense, and presupposition.

An example of this work in mobile language applications [74] uses short messaging services, smartphone apps, and gamification to enhance parental behavior that promotes language development in children. The work in this specific example is geared more toward parents who can implement...
interventions for their children; however, configurations for other neurocognitive conditions can also be formed.

**Executive Function**

**Background and Subcategories**

Executive function refers to the abilities of judgment, planning, memory, efficiency, and time management and is relevant in the assessment of neurocognitive functioning and decline. A decline in executive function can be seen in neurodegenerative conditions [28,75], neurodevelopmental conditions [32], and TBI and/or ABIs [76]. Similar to memory analysis, executive function can fluctuate due to factors including stress and fatigue; however, constant decline trends in executive function can be used as an indicator for disease diagnosis. Traditional testing modes for the purpose of executive function are referenced in Tables 3 and 8.

<table>
<thead>
<tr>
<th>Executive function subcategories and functional tests</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Judgment</strong></td>
<td></td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>[26]</td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>[20]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>[22]</td>
</tr>
<tr>
<td><strong>Planning</strong></td>
<td></td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>[23]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>[22]</td>
</tr>
<tr>
<td><strong>Time management</strong></td>
<td></td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>[26]</td>
</tr>
<tr>
<td>Wisconsin Card Sorting Test</td>
<td>[22]</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td></td>
</tr>
<tr>
<td>Bender-Gestalt Test</td>
<td>[24]</td>
</tr>
<tr>
<td>Spatial Orientation Tests</td>
<td>[25]</td>
</tr>
<tr>
<td>Stroop Color Word Test</td>
<td>[20]</td>
</tr>
<tr>
<td>Trail Making Test</td>
<td>[23]</td>
</tr>
</tbody>
</table>

**Mobile Assessments**

Analysis of executive functions implements human-device interactions [50] in conjunction with positional [37,38,41], media [42], and inherent [49] device sensors, as seen in Table 4. Similar to motor function, human-device interactions for the purpose of executive function utilize sensors to monitor the positional state of a user’s finger on the screen (eg, electrical current or reflection of waves). The output of this can be expressed as coordinates in 2-dimensional space [52], which provides opportunities for the enhancement of planning and efficiency tests (eg, the Trail Making Test). Positional sensors for the capture device motion [37,38] can be used in the enhancement of object manipulation tests (eg, having the subject manipulate the mobile device itself). GPS positional sensors can be used by having a person go from one place to another and seeing how long it takes them and the route they take [41]. Media sensors, with the purpose of speech recognition in enhancing the Boston Naming Test, can be used to collect accuracy metrics in the subject’s discernment of images [42]. Finally, the device timer is highly important in executive function analysis as it yields temporal metrics for the purpose of time management, efficiency, and judgment [49].

Current monitoring of executive function on mobile devices uses concrete tests from clinical practice, such as the Trail Making Test [77], and more abstract tasks such as prioritization and planning in scheduling [78]. Both methods are configurable for mobile apps; however, different timelines and tracking metrics are given as outcomes.

**Sensory**

**Background and Subcategories**

Evaluation of visual, tactile, and aural senses are important in neurocognitive assessments, as these senses can be affected by neurocognitive conditions. Autonomic dysfunctions, including dizziness, sensation, and blurred vision, are all reasons that these sensory components should be monitored for an all-inclusive neurocognitive assessment. In addition, an individual’s perception of pain, or lack of feeling, is another sensory metric that is important to assess in neurodegenerative conditions [79] or TBI and/or ABIs [80]. A reference for the traditional assessment of sensory function is given in Table 3.

**Mobile Assessments**

Device capabilities for the monitoring of sensory function are shown in Table 4. Reaction to visual, tactile, or aural stimuli through participatory device interactions [51] or metadata input [52] regarding sensory functions are common modes for the analysis of this functional section. Mobile implementations for these stimulus responses can be implemented using vibrational...
patterns, screen display changes, or auditory sounds configured on the device. Screen interactions or device sensors can then be used to gauge user feedback on the signals based on configurations. Redesigned metadata surveys and questionnaires for the collection of user data allow the depiction of their sensory state. Mobile devices can allow more real-time reporting of these symptoms by allowing the user to label their pain levels throughout the day in conjunction with information on their current state (eg, during rehabilitation or right after sleep).

Current sensory function monitoring on mobile apps is both quantitative [60] and qualitative [81] in nature. A study [60] used visual, tactile, and aural feedback in conjunction with gait rehabilitation and training. This research evaluated the influence of sensory feedback on the gait pattern of the subject in real time, for the purpose of clinical rehabilitation of persons with gait abnormalities [60]. Qualitatively, self-management systems are used in practice to assist in rehabilitation by supporting goal setting and providing user state information and feedback [81].

**Behavioral and Psychological**

**Background and Subcategories**

Behavioral and psychological function assessment is necessary, as neurocognitive conditions portray emotional changes after onset. This could present with the inability to express or understand different emotions [82] or expose changes to the person’s outlook [29]. This is relevant to assessing multiple capacities as *emotion* is an important feature of social interactions and quality of life and well-being. A reference in Table 3 is given for the traditional assessment of behavioral and psychological function.

**Mobile Assessments**

References on the completion of behavioral and psychological monitoring using a device’s media (eg, camera and microphone) sensors [44,45,47] as well as metadata input [52,53], device lock and unlock, and app usage data [54] are provided in Table 4. Emotional state assessment can be completed using device media sensors through the analysis of speech and/or video samples [44,45,47]. Processing these samples, using machine learning approaches, can assist with the classification of the emotional state of its users. Similarly, the configuration of device labeling allows the user to provide a more real-time depiction of their state of being throughout the day [52,53]. The use of metadata inputs can also help with the monitoring of medication cycles and/or interpersonal relations in conjunction with mood or emotional behavior. Finally, metrics on device lock and unlock and app usage provide viable information for the emotional state. User reliance on technology and its correlation with interpersonal connections are relevant to monitor in conjunction with the emotional state of the user. Collecting these metrics directly from the user’s phone to an assessment app makes for an overall smart system.

According to Pavliscak et al [83], mobile health apps for the collection of information regarding behavioral and psychological states are highly useful and successful, in addition to standard care measures through increased interactions. Mobile app questionnaires about user health status, psychosocial status, and progress toward treatment goals were implemented. Similarly, Juegst et al [84] explored the use of mobile apps for mood-related symptom tracking post TBI. Both studies looked at compliance, satisfaction, and usability measures for the validation of apps in practice. All metrics yielded high values, supporting the collection of this information via a smartphone.

**Sleep**

**Background and Subcategories**

There are direct correlations between sleep abnormalities and neurocognitive diseases and conditions, making sleep a valuable component for neurocognitive analysis. This relationship occurs for all categories: neurodegenerative, neurodevelopmental, and neuropsychological conditions as well as TBIs or ABIs [36,85-87]. Individuals who have abnormalities in their sleep patterns ultimately show additional abnormalities among other functions [36].

**Mobile Assessments**

The sleep-monitoring capabilities that mobile devices contain are shown in Table 4. Positional sensors for movement in sleep [39,40], media sensors for sleep apnea [40], and the device timer for duration of sleep [49] are all helpful in monitoring sleep quality. Metadata input [52], lock and unlock, and app usage metrics [54,55] are also necessary for monitoring sleep quality. In sleep analysis, positional sensors (eg, accelerometer and gyroscope), measure device and subsequent user motion, or lack thereof, for assessment metrics. Microphone usage for breathing patterns is helpful for the monitoring of sleep apneas. Device timers in conjunction with both allow the temporal analysis of sleeping patterns for the evaluation of sleep. Similar to emotion, metrics on device lock and unlock and app usage provide viable information for sleep assessments, as user reliance on technology may have a negative correlation with sleep patterns [54]. Configured metadata input (eg, labeling) from the user can allow for consistent monitoring of their sleep over time, providing historical monitoring of sleep quality and quantity.

There are many current mobile apps for sleep monitoring and analysis [88]. These apps range in functionality but track total sleep time, duration of light or deep sleep, and time awake [88]. A study [89] used explicit interaction of the subject with a mobile app to monitor sleep duration. App functionalities include an alarm, labeling functionalities for sleep versus awake, and a rating system to gauge sleep quality [89]. Monitoring of users’ sleep behavior is done through the logging of metrics including: set alarm time, scheduled wake up time, time of day in which the user goes to bed, number of times the alarm is snoozed, duration of the snooze, and time of day when the alarm is deactivated. This study [89] suggests that providing more methods for users to track sleep behaviors increased the awareness of their sleep patterns and induced healthier habits.

**Autonomic Function**

**Background and Subcategories**

Autonomic functions are processes that the body regulates unconsciously (eg, heart rate, respiration, swallowing, thermal regulation, digestion, and pupillary response). These functions may be affected by the onset of neurological conditions [90].
but may be the result of drug therapy side effects [91]. A reference for the traditional assessment of autonomic function is given in Table 3.

Mobile Assessments

References on device capabilities for the monitoring of autonomic functions are depicted in Table 4. Device media sensors (eg, microphone [40] and camera [48]) are useful for monitoring functions such as breathing and pupillary response. Metadata input [49] is helpful regarding other autonomic functions (eg, digestion or urination) and may be relevant for drug intervention analysis. Although some mobile phones have heart rate sensors; Table 4 is a representation of functionalities and sensors that most mobile phones contain. Sound and image sample processing techniques (eg, machine learning) can be implemented on these devices for gaining metrics of the user’s autonomic state. Metadata input for the collection of additional autonomic functional information, which cannot be collected by device sensors, allows for a more comprehensive assessment of this area.

Current mobile apps for autonomic function monitoring include the evaluation of both breathing [92] and heart rate [93]. In a study [92], 3 training methods were created to see which provided the best outcomes. To establish which breathing training method worked best, formal metrics were collected in the following areas: skin conductance, heart rate, and respiratory signal-to-noise ratio, whereas perceived effectiveness and subjective preference were collected using questionnaires. Current work for in-home monitoring of acute and chronic cardiovascular disease uses mobile devices for both the collection of heart rate and physical activity data sent to a mobile phone via Bluetooth [93]. The mobile phone app is then used for analysis and long-term storage of information to measure progress and can be viewed by both the subject and clinician.

Discussion

Future of Mobile Neurocognitive Assessments

Devices

Current device capabilities can and should be explored for the future of neurocognitive assessments. Employing opportunistic approaches to monitoring (having device sensors on in the background without the need for formalized tests) allows for additional collection methods of objective data. An example of this approach would use the device’s GPS sensors in the background to gather information on daily commutes to see if patterns change over time. Understanding device limitations is another important aspect in this area, as data on the device cannot be collected endlessly. Participatory, opportunistic, and even hybridized approaches; further employment of current device capabilities, collection of objective data (eg, sensor metrics), and collection of additional metadata from the user, should all be addressed for the formation of an all-encompassing neurocognitive assessment. These mobile devices need to allow for additional wearable and/or internet of things (IoT) devices to interact with one another. Data fusion approaches, maintaining overall battery and data usage on devices, protecting user privacy, among others, are areas of concern that are important for device advancement and the future of these mobile neurocognitive assessments.

User Interactions

Moving neurocognitive assessments to mobile platforms for users allows them to explore, understand, and maintain another facet of their overall health. The ability to directly interact with their devices for training exercises, neurocognitive assessments, or rehabilitation purposes with regard to neurocognitive function allows users to have a sense of control and ownership of an important aspect of their health. Possessing these assessments on their mobile device affords users the ability to track their progress and see relevant longitudinal data. The functionality of these mobile devices is intended to not only make the user feel in control but also give the user paramount tools to assess their neurocognitive function compared with previous clinical versions. Concerns of the device to be addressed in the transition to mobile devices include preserving the privacy of their personal information and maintaining data and/or battery usage on their devices, while having positive and simple interactions for the assessment. These simple interactions require foresight in the creation of mobile testing versions.

Clinician Interactions

As neurocognitive testing becomes readily available on mobile devices, it is important to maintain clinical expertise. Clinical challenges arise, such as how the user interprets instructions and possible data quality and consistency issues (eg, in the cases of different neurological states between healthy populations and diagnosed neurologically impaired populations, test-retest problems, language barriers, or others). Similarly, when moving clinical assessments to mobile devices for additional sensor data, it is important to maintain relevant metadata on how the user feels and interprets their own symptoms, as there may be fewer interactions with clinical professionals who would administer questions, evaluate, and observe users. The clinician should use these devices to monitor the user, analyze the respective objective data, and ultimately assist in diagnosing conditions and formulating rehabilitation programs, if necessary. Concerns of the clinician include mobile device users diagnosing conditions on their own, a large influx of overall data, as well as maintaining the user’s personal information and the patient and clinician relationship. The benefits of these systems include the clinician’s prior review of objective and concise data, such that they can spend more time talking with the patient about specific or personal issues regarding their disease.

Wearables

Wearables and other functional sensing systems that work in conjunction with mobile devices can allow for even more vital data to be collected. Devices include, but are not limited to, smart watches or necklaces, fitness trackers, and even implantables. Wearables can be used in conjunction with mobile devices, or even separately, and both methods have their benefits and challenges. With the implementation of wearable devices into the system, an enhanced set of data can be obtained in addition to new information that the mobile device may not be able to detect on its own. This is directly related to more health-related sensors such as heart rate and oxygen saturation.
These data are important to be collected continually to monitor a user’s current state, exertion levels, etc.

Another benefit of wearables is the ability to obtain more data with additional accelerometers and gyroscopes. These additional sensors can allow for the collection of imperative data throughout the day, enhancing neurocognitive assessment systems. However, with this additional data from integrated wearables, there is a need to merge the collected data, specifically in the case of accelerometer and gyroscope use. Data fusion can be completed in a variety of ways, and each functional task might call for different fusion methodologies. For example, certain functional tasks such as motor function, including gait, balance, and sit-to-stand tasks, would be inherently beneficial for assessing with both mobile devices and wearables in tandem to get a more complete look as to how the individual moves in space. This can be seen by monitoring both devices’ positioning in space, thus providing a proximity component to the analysis. Other motor functional tests, however, such as fine motor skills and some reflex tests, may not have much of a response on the smartphone device depending on how the user interacts with the test (eg, if the device is lying on a table while being interacted with, the wearable becomes the primary source of data collection). It is imperative that data are collected on all devices when in active use; however, one device’s dataset could provide significant insights for certain tests. Subsequently, data fusion will be an area of focus when multiple devices are implemented in the same system.

Overall Challenges and Opportunities

As traditional assessments move to mobile devices, multiple challenges arise that need to be considered and addressed. Challenges can occur within each functional area of assessment (eg, motor, memory, and speech). The monitoring of each functional area or respective subfunction requires unique configurations of a variety of device sensors. Each disease taxonomy could call for unique configurations. For example, children would require much different device interactions than older populations (eg, neurodevelopmental vs neurodegenerative conditions). Testing instructions (eg, size for visualization, lay language styles, and memory restrictions) pose challenges for device assessments. The formation of quality apps that are both detailed and understandable is important for both users and caregivers (as there are subsets of users with neurocognitive conditions that cannot complete these tasks on their own). In addition, as these devices are to be used outside of clinical settings, sample quality (eg, image or sound) poses challenges with lighting and background noise. This requires either an isolated environment to remove potential noise or filtering methods based on these files. In addition, distinguishing when to use a certain collection or assessment method over another across functional areas or combining neurocognitive functions for multimodal analysis remains a challenge. As certain device sensors are used across multiple functional areas, multimodal tasks are achievable (eg, The Stroop Color Word Test for both judgment and speech data), which reduce the administration time for functional analysis. The design of these tests, however, becomes more intensive as more metrics need to be collected. In addition, there are multiple functions that occur in ways that are, unfortunately, not easily monitorable by standalone mobile devices (eg, sleep as the user may not have a smartphone on their person or digestion as this process happens in a way that is not monitorable by a smartphone’s device sensors). Monitoring these functions requires the use of more inclusive IoT systems (eg, using smart-home technologies or other monitoring devices such as wearables). Other wearable opportunities include increasing the monitoring and real-time analysis of important features (eg, heart rate) and inclusion of new features (eg, galvanic skin response, temperature regulation, and pulse oximetry).

The collection of more objective data metrics is highly beneficial for both users and clinicians. Subjective biases are reduced with the implementation of these new impartial metrics. With the increased opportunity for user participation in their own devices and the ability of the clinician to collect and analyze enhanced objective datasets, this becomes a robust modality for the administration of these neurocognitive assessments. The use of mobile devices for assessment allows for more continual fine-grained monitoring and historical comparisons.

According to Furlong et al [69], there are few (approximately 3%) apps that are therapeutically beneficial for respective function monitoring. Similarly, there are numerous health apps in the app store that can measure some of the functions of interest; however, no apps measure all functions [3]. Highly specific apps for monitoring certain conditions are objectively helpful. Robust general apps, however, should be created for monitoring individuals before diagnosis. These general apps should be more than just screening tools before additional testing, but rather comprehensive apps. The formation of multiple monitoring and testing techniques should be completed for effectiveness comparisons. This would allow for highly standardized comprehensive assessment suites that can then feed into specific apps when necessary (eg, postdiagnosis or unique user conditions). Finally, although there are both notable challenges and opportunities proposed in this work, there are likely additional concerns that are not discussed, but equally important. As the relationship between mobile devices and health care deepens, the lists of challenges and opportunities will likely grow in tandem.

Conclusions

The relationship between mobile devices and health care for the purpose of neurocognitive assessment is underway; however, due to the area being relatively young and the expansive possibilities of mobile technology, there are still numerous new avenues to be explored and/or enhanced. Upgrading mobile technology for these assessments and employing inherent device capabilities and human interactions will ultimately allow for a deeper understanding of neurological diseases. Configurations of current mobile sensors, new assessment approaches, addition of new sensors into the system, new expansive IoT systems, and exploration of data fusion and deep learning techniques for these assessments are all ways to further this adolescent connection between health care and mobile devices, not only to augment clinical interactions with users’ devices but also the overall purpose of objective and comprehensive neurocognitive assessments.
Conflicts of Interest

None declared.

References


http://mhealth.jmir.org/2020/6/e15517/


http://mhealth.jmir.org/2020/6/e15517/


78. Temple C. Perspectives on Language and Literacy. 2013 Presented at: IEMBS'09; September 3-6, 2009; Minneapolis, MN, USA. [doi: 10.1109/iembs.2009.5333540]


**Abbreviations**

- ABIs: acquired brain injuries
- TBIs: traumatic brain injuries
- IoT: internet of things
©John Michael Templeton, Christian Poellabauer, Sandra Schneider. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org). 24.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Feasibility and Utility of mHealth for the Remote Monitoring of Parkinson Disease: Ancillary Study of the PD_manager Randomized Controlled Trial

Dimitris Gatsios\textsuperscript{1,2}, BSIT; Angelo Antonini\textsuperscript{3,4}, MD, PhD; Giovanni Gentile\textsuperscript{3,4}, MSc; Andrea Marcante\textsuperscript{4}, MD, PhD; Clelia Pellicano\textsuperscript{5}, MD, PhD; Lucia Macchiusi\textsuperscript{5}, PhD; Francesca Assogna\textsuperscript{5}, PhD; Gianfranco Spalletta\textsuperscript{5}, MD, PhD; Heather Gage\textsuperscript{6}, PhD; Morro Tournay\textsuperscript{6}, PhD; Lada Timotijevic\textsuperscript{7}, PhD; Charo Hodgkins\textsuperscript{7}, PhD; Maria Chondrogiorgi\textsuperscript{1}, MD; George Rigas\textsuperscript{2}, PhD; Dimitrios I Fotiadis\textsuperscript{2,8}, PhD; Spyridon Konitsiotis\textsuperscript{1}, MD, PhD

\textsuperscript{1}Department of Neurology, Medical School, University of Ioannina, Ioannina, Greece
\textsuperscript{2}Unit of Medical Technology and Intelligent Information System, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece
\textsuperscript{3}Department of Neuroscience, University of Padua, Padua, Italy
\textsuperscript{4}San Camillo Hospital Istituto Di Ricovero e Cura a Carattere Scientifico, Venice, Italy
\textsuperscript{5}Laboratory of Neuropsychiatry, Fondazione Santa Lucia Istituto Di Ricovero e Cura a Carattere Scientifico, Rome, Italy
\textsuperscript{6}Surrey Health Economics Centre, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom
\textsuperscript{7}School of Psychology, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom
\textsuperscript{8}Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology-Hellas, Ioannina, Greece

Corresponding Author:
Dimitris Gatsios, BSIT
Department of Neurology
Medical School
University of Ioannina
University Campus of Ioannina
Ioannina, 45110
Greece
Phone: 30 2651007261
Email: d.gatsios@uoi.gr

Abstract

Background: Mobile health, predominantly wearable technology and mobile apps, have been considered in Parkinson disease to provide valuable ecological data between face-to-face visits and improve monitoring of motor symptoms remotely.

Objective: We explored the feasibility of using a technology-based mHealth platform comprising a smartphone in combination with a smartwatch and a pair of smart insoles, described in this study as the PD_manager system, to collect clinically meaningful data. We also explored outcomes and disease-related factors that are important determinants to establish feasibility. Finally, we further validated a tremor evaluation method with data collected while patients performed their daily activities.

Methods: PD_manager trial was an open-label parallel group randomized study. The mHealth platform consists of a wristband, a pair of sensor insoles, a smartphone (with dedicated mobile Android apps) and a knowledge platform serving as the cloud backend. Compliance was assessed with statistical analysis and the factors affecting it using appropriate regression analysis. The correlation of the scores of our previous algorithm for tremor evaluation and the respective Unified Parkinson’s Disease Rating Scale estimations by clinicians were explored.

Results: Of the 75 study participants, 65 (87\%) completed the protocol. They used the PD_manager system for a median 11.57 (SD 3.15) days. Regression analysis suggests that the main factor associated with high use was caregivers’ burden. Motor Aspects of Experiences of Daily Living and patients’ self-rated health status also influence the system's use. Our algorithm provided clinically meaningful data for the detection and evaluation of tremor.

Conclusions: We found that PD patients, regardless of their demographics and disease characteristics, used the system for 11 to 14 days. The study further supports that mHealth can be an effective tool for the ecologically valid, passive, unobtrusive monitoring of Parkinson disease.

https://mhealth.jmir.org/2020/6/e16414

JMIR Mhealth Uhealth 2020 | vol. 8 | iss. 6 | e16414 | p.255
(page number not for citation purposes)
monitoring and evaluation of symptoms. Future studies will be required to demonstrate that an mHealth platform can improve disease management and care.

**Trial Registration:** ISRCTN Registry ISRCTN17396879; http://www.isrctn.com/ISRCTN17396879

**International Registered Report Identifier (IRRID):** RR2-10.1186/s13063-018-2767-4

*(JMIR Mhealth Uhealth 2020;8(6):e16414)* doi:10.2196/16414

**KEYWORDS**
Parkinson's disease; determinants of compliance; clinically meaningful data; ecological validity

**Introduction**

Parkinson disease (PD) is a progressive disorder with patients having heterogeneous symptoms and progression rates. Presently, there is no cure for the condition, and treatment aims at controlling symptoms by optimizing medication plans. Optimization and personalization of the treatment is currently based on clinical interview, diaries, and scales, although in the future it may benefit from information on symptoms and medication adherence collected away from the clinic while the patients perform their normal daily activities.

Technology, even with the existing regulatory limitations and barriers, offers the possibility for improved care, self-assessment options, and overall improved health care outcomes [1]. Wearable sensors and mobile apps have been extensively used to monitor and evaluate mainly motor symptoms and motoric complications of PD patients in their home environments [2]. However, reliable and unobtrusive solutions for nonmotor symptoms are still lacking [1].

Despite the potential benefits of the use of technologies, important aspects of its feasibility remain to be explored. Only a few studies have rigorously investigated the feasibility and utility of using technology-based platforms. Moreover, apart from three studies [3-5], most prior studies remained limited by the small sample sizes (samples of up to 51 PD patients in varying disease stages) [6-11]. Evidence of mHealth utility for the clinicians is in its early days even for commercial grade systems [12,13]. None of the previous studies has systematically explored the role of caregivers in compliance with mHealth.

In this analysis, we aimed to investigate the feasibility of using an mHealth platform, described in this study as the PD_manager system, comprising a smartphone, smartwatch, and pair of smart insoles. The study focuses on participants’ compliance and their determinants. The study also validates the system’s utility to collect clinically meaningful data with ecological validity.

**Methods**

**Study Population**

Between May 2017 and March 2018, 136 consenting patients with PD (Hoehn and Yahr scale stage of ≥3, experiencing motor fluctuations at least 2 hours per day based on Unified Parkinson’s Disease Rating Scale [UPDRS] IV score), with a live-in caregiver, were recruited in three countries (50 in Rome and 44 in Venice, Italy; 21 in Ioannina, Greece; and 21 in Surrey, England). Four of them were excluded from the study, 2 because they withdrew and 2 because they were not eligible at reassessment, leaving a total of 75 patients assigned to the PD_manager group and 57 to the control group. The PD_manager group characteristics are summarized in Table 1.

<table>
<thead>
<tr>
<th>Table 1. PD_manager group characteristics (n=75).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Gender (female), n (%)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
</tr>
<tr>
<td>Years since diagnosis, mean (SD)</td>
</tr>
<tr>
<td>H&amp;Ya (stage 3 patients), n (%)</td>
</tr>
<tr>
<td>BMI (%), mean (SD)</td>
</tr>
<tr>
<td>MMSEb, mean (SD)</td>
</tr>
<tr>
<td>UPDRS IIIc, mean (SD)</td>
</tr>
<tr>
<td>NMSSd, mean (SD)</td>
</tr>
</tbody>
</table>

| aH&Y: Hoehn and Yahr scale.                  |
| bMMSE: Mini-Mental State Examination.        |
| cUPDRS III: Unified Parkinson’s Disease Rating Scale III. |
| dNMSS: Nonmotor Symptoms Scale.             |

In this work, we were focusing on actual system use and compliance as depicted in the data collected from the patients and their caregivers vis-à-vis data collected via the respective PD_manager devices (ie, we were analyzing only the
PD_manager group). Focus was on compliance metrics as well as the factors affecting the compliance. Moreover, we provide evidence that the data are clinically meaningful since they can be used for accurately monitoring and evaluating symptoms and specifically tremor.

**Study Design**
The PD_manager trial [14] was an open-label parallel group randomized study. It was conducted to assess the feasibility, usability and trends of effectiveness of the PD_manager system compared with traditional practices of using a symptom diary for the management of people with PD.

Following informed consent, baseline information was gathered, including the following: age, gender, education, attitudes toward technology (patient and caregiver), time since Parkinson diagnosis, symptom status (with Nonmotor Symptoms Scale (NMSS) and UPDRS), comorbidities (patient only), caregiver burden (Short Zarit), patient’s self-assessment of the disease (EuroQol 5-Dimension 5-Level [EQ-5D-5L]) and patient’s self-assessed quality of life (Parkinson’s Disease Questionnaire–8).

Patients were asked to use the system for 14 days continuously for 12 hours during the day. The 14-day duration for the wearing of study devices (wristband and smartphone) by participants was selected for a number of reasons. First, it was based on analysis of user needs, safeguarding ethics and privacy, as well as the burden on study participants. Second, it was considered enough for collection of sufficient data to provide clinically meaningful information. Finally, findings of previous larger studies [3], with similar investigation concepts, indicated that around 70% of the patients were compliant for up to 15 days.

During the 14-day period, the system passively and automatically captured raw sensor data (from the smartphone, wristband, and insoles) to be used for the evaluation of motor symptoms, aggregated data on sleep and activity (wristband proprietary software), speech, cognitive status, and emotional state using the smartphone apps (with scheduled prompts for the user to perform specific tasks). The smartphone was used for storing the data locally. Automatic transmission of the data to a cloud backend was possible but not used during the pilot for privacy and security purposes. Control group participants were asked to keep a motor symptom diary for 3 days and complete the Parkinson Well-Being Map. After a minimum of 2 weeks, a specialist doctor reviewed the data gathered. Participants, caregivers, and clinicians were asked for feedback on the acceptability and utility of the data collection methods. Data collection for the pilot study is summarized in **Table 2**.

### Table 2. Summary of PD_manager group data collection at each stage.

<table>
<thead>
<tr>
<th>Participant group</th>
<th>Data capture at each stage</th>
<th>During intervention, PD_manager group from devices</th>
<th>Postintervention, 2-week follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
<td>Age, gender, education, disease duration, disease stage (Hoehn and Yahr score), main symptoms (tremor, bradykinesia, rigidity, dyskinesia), more affected side, UPDRS scores, current medications, comorbidities, views on technology (with the TAMM). Outcomes: EQ-5D-5L, PDQ-8, NMSS, UPDRS</td>
<td>Motor symptoms (gait, freezing of gait, bradykinesia, dyskinesia, activity); nonmotor symptoms (cognition, sleep, mood)</td>
<td>Interviews on acceptability and ease of use of PD_manager or symptom diary; Data collected in the smartphone and in the backend from smartphone and wristband sensors, data from insoles stored in the backend</td>
</tr>
<tr>
<td>Caregiver</td>
<td>Age, gender, education, views on technology (with the TAMM), Outcome: Zarit Caregiver Burden Scale (short version)</td>
<td>No information was collected from caregivers in the PD_manager group</td>
<td>Interviews on acceptability and ease of use of PD_manager or symptom diary</td>
</tr>
<tr>
<td>Clinician</td>
<td>Technophobia, previous experience with monitoring technology, sociodemographics, clinical experience</td>
<td>—</td>
<td>SUS, PSSUQ, TAMM.</td>
</tr>
</tbody>
</table>

*UPDRS: Unified Parkinson’s Disease Rating Scale.
*TAMM: technology acceptance modified model.
*EQ-5D-5L: EuroQol 5-Dimension 5-Level.
*PDQ-8: Parkinson’s Disease Questionnaire–8.
*NMSS: Nonmotor Symptoms Scale.
*SUS: system usability scale.
*PSSUQ: Poststudy System Usability Questionnaire.

**mHealth Platform**
The mHealth platform depicted in Figure 1 has been described in detail in a previous work [15] and consists of a wristband (Microsoft Band, Microsoft Corporation), a pair of sensor insoles (Moticon GmbH), a smartphone (Aquaris M and U models, BQ) with dedicated mobile Android apps (see Figure 2) and a knowledge platform (hosted by Biotronics 3D) serving as the cloud backend of the platform. The Microsoft Band software development kit allowed us to access data from the Band’s sensors. The wristband and smartphone provided raw data from the 3-axis accelerometer and gyroscope at a sampling rate of 100 Hz that were used for building motor symptoms’ assessment methods. The Band could also be used for collecting heart rate, galvanic skin response, and skin temperature data. Moreover, the accompanying Microsoft Health App provided...
aggregated data for sleep (sleep duration, number of wakeups, ratio of time asleep to total sleep, total length of restless, and restful sleep in minutes) and activity type (e.g., run, sleep, bike, summary of calories burned, summary of heart rate data). With the insoles, we collected pressure distribution and accelerometer data enabling us to evaluate weight-bearing, balance and motion sequences, and study gait.

Figure 1. PD_manager mHealth platform overview.

Figure 2. Screenshots from mobile app. From left to right: tasks patient must perform, some cognitive tests, finger-tapping test, mood monitoring diary.

Data from devices were transferred and stored in a web-based cloud, NoSQL database in anonymized and encrypted format. The servers storing the information in the cloud platform are based on Biotronics 3D’s 3DnetMedical platform in an ISO 27001-accredited data center located in London. They are operated in accordance with the Data Protection Act. During the pilot study, participants were instructed to always carry the smartphone with them since the wristband needs to be paired with the phone through the Bluetooth connection for transmitting and storing wristband data. Two optimization strategies were applied in order to reach the desired Microsoft Band battery daily duration of 12 to 14 hours and address battery

https://mhealth.jmir.org/2020/6/e16414
drain issues: (1) by default, the app acquires data for a period of 5 minutes and then disconnects from Microsoft Band (closing Bluetooth and therefore significantly reducing the Microsoft Band power consumption) for a period of X minutes, where X is estimated based on the hours of the required recording interval which is customized in the app settings, and (2) when the patient removes the Microsoft Band (detected with the heart rate quality value), the data acquisition is postponed. Moreover, study participants were instructed to use the system as much as possible during the waking day while performing daily activities and charge it just before going to sleep. The insoles had their internal storage capability.

The devices are unobtrusive. Their wearability, sensitivity, and reliability were tested as part of an earlier proof of concept study [16] with 20 patients (5 Rome and 10 Venice, Italy; 5 Ioannina, Greece). This proof-of-concept study was supervised by neurologists in an in-hospital setting and involved short sessions (154 in total, each lasting around 30 minutes) following a common protocol that included simulation of daily activities such as opening a door, drinking water, walking a few meters, rising from a chair, and rising from the bed. The nutrition and physiotherapy modules were evaluated in separate studies [17].

The clinicians had a dedicated mobile app (see Figure 3) that enabled them to check the demographic and clinical information, assess the overall status of the patient, evaluate symptoms monitored during the pilot period, and get decision support functionalities [18] on patient mobility.

Figure 3. Screenshots from clinician mobile app. From left to right: overview of clinical information, scores from scales and tests, overview of motor symptoms as assessed by the PD_manager, and create new medication order. To ensure there are no risks for participants, we omitted the medication adherence module (mobile app and pillbox) from pilot.

Outcome Definitions and Statistical Analysis
Feasibility assessment included recruitment, compliance, and evaluation of the processed sensor data utility for answering clinically meaningful questions. Recruitment success was analyzed by the total number of enrolled, consenting participants who completed the pilot study against dropouts. Compliance was calculated as the total hours where band and smartphone sensor data were collected during the 14-day period, as well as number of days during which the participant used the system for at least 1 hour.

The statistical analysis investigated the effects of the patient demographics (age, gender, education), clinical symptoms (as depicted in the Nonmotor Symptom Scale and UPDRS), self-rated quality of life (Parkinson’s Disease Questionnaire–8 (PDQ-8) and EQ-5D-5L), caregivers’ demographics (age, gender, education), and burden (as captured with the short version of Zarit) on the system use as reflected in the total use hours over the 14-day data collection period by the devices for each participant. In this targeted analysis, we have included only the 65 of the originally recruited 75 participants for which duration of data collected is at least one day of the pilot period.

The study data were analyzed by SPSS Statistics version 23 software (IBM Corporation).

Compliance was not normally distributed. Correlations between compliance and the available at-baseline information for the participants were explored with Spearman rank-order and Kendall tau-b. Participants were then divided in low, moderate, and high use groups using the quartiles (the first quartile was the cutoff for the low compliant group and third quartile for the high compliant group) and taking into account qualitative information, mainly Band use—another metric available for compliance evaluation—for confirming the grouping. Significant differences in the distributions of use between compliance groups were investigated with a Kruskal-Wallis H test for the low, moderate, and high groups.

To further investigate factors affecting compliance, regression analysis was applied. Linear regression determined how much of the variation in the use was explained by the caregiver burden. Multiple linear regression determined how much of the variation in the system use was explained by the caregivers’ burden, Motor Aspects of Experiences of Daily Living, and patient self-rated health statuses. Binary logistic regression explored the effects of the same parameters on the likelihood of use, predicting the moderate and high groups.
Validation of the tremor method was done with bivariate correlations (with Pearson test) between UPDRS items scored by the clinicians at baseline, and the tremor score calculated with our method. A Welch $t$ test was also run to determine if there were differences in scores between the no-tremor and tremor groups. The statistical methods used in the analysis are depicted in Figure 4.

**Figure 4.** Outcome definitions and statistical analysis.

### Results

#### Recruitment

From the 75 patients who were eligible and consented to participate in the study and were randomly assigned to the PD_manager group, 65 (87%) were data providers with at least 1 day of system use. The other 10 either chose not to use the system or due to technical reasons (Bluetooth disconnection) were unable to use it.

#### Compliance

The 65 data contributors collected data for a median of 63.37 (SD 42.17) hours total in the 14-day study period (ie, 4.53 hours on average per day). They used the system for a median 11.57 (SD 3.15) days. Only two of the study participants used the system for 1 day. All others used it for more than 6 days, with 30 using it for the whole 14-day study period.

#### Sample Characteristics and Bivariate Correlations

Study sample characteristics are presented in Table 3. Most participants had many symptoms as reflected in UPDRS total and subscores. Most study participants were men (almost 2:1), while for caregivers the reverse was observed (women 2:1). Caregivers were slightly more educated (11.96 years) compared with study participants (10.18 years), which can be explained by the fact that 27% were children or nephews. For the same reason, the caregivers were younger (mean 60 years). UPDRS score (mean 56.45) is consistent with the severity of the condition. Participants did not have dementia (based on Mini-Mental State Examination and instrumental activities of daily life).
### Table 3. Analysis of distributions between groups.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>High group (n=21)</th>
<th>Moderate group (n=28)</th>
<th>Low group (n=16)</th>
<th>$\chi^2$ (P value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Correlation coefficient, $P$ value (spearman)</td>
<td>Mean (SD)</td>
<td>Correlation coefficient, $P$ value (spearman)</td>
</tr>
<tr>
<td><strong>Use hours</strong></td>
<td>108.29 (37.50)</td>
<td>.10 (.66)</td>
<td>58.93 (18.47)</td>
<td>-.03 (.88)</td>
</tr>
<tr>
<td><strong>Days of use</strong></td>
<td>14.8 (.001)</td>
<td>.34 (.20)</td>
<td>9.63 (4.40)</td>
<td>-.15 (.64)</td>
</tr>
<tr>
<td><strong>Caregiver age</strong></td>
<td>60.00 (13.09)</td>
<td>.19 (.43)</td>
<td>58.38 (12.06)</td>
<td>.17 (.40)</td>
</tr>
<tr>
<td><strong>Caregiver education</strong></td>
<td>10.89 (4.90)</td>
<td>-.26 (.28)</td>
<td>12.50 (5.02)</td>
<td>.13 (.54)</td>
</tr>
<tr>
<td><strong>Patient age</strong></td>
<td>67.24 (6.71)</td>
<td>-.23 (.31)</td>
<td>67.67 (11.45)</td>
<td>.19 (.35)</td>
</tr>
<tr>
<td><strong>Patient education</strong></td>
<td>9.50 (4.87)</td>
<td>-.42 (.07)</td>
<td>10.48 (4.23)</td>
<td>.002 (.99)</td>
</tr>
<tr>
<td><strong>Disease duration</strong></td>
<td>8.90 (5.04)</td>
<td>.04 (.85)</td>
<td>9.18 (4.67)</td>
<td>-.08 (.69)</td>
</tr>
<tr>
<td><strong>MMSE</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>28.32 (1.70)</td>
<td>.58 (.01)</td>
<td>28.79 (2.13)</td>
<td>.17 (.38)</td>
</tr>
<tr>
<td><strong>EQ-5D-5L</strong>&lt;sup&gt;b&lt;/sup&gt; total</td>
<td>9.48 (3.63)</td>
<td>-.20 (.38)</td>
<td>10.54 (3.42)</td>
<td>.06 (.78)</td>
</tr>
<tr>
<td><strong>NMSS</strong>&lt;sup&gt;c&lt;/sup&gt; total</td>
<td>42.09 (29.43)</td>
<td>-.16 (.48)</td>
<td>44.00 (34.85)</td>
<td>.21 (.29)</td>
</tr>
<tr>
<td><strong>PDQ-8</strong>&lt;sup&gt;d&lt;/sup&gt; total</td>
<td>41.25 (18.41)</td>
<td>-.10 (.68)</td>
<td>43.29 (23.95)</td>
<td>.19 (.34)</td>
</tr>
<tr>
<td>**UPDRS&lt;sup&gt;e&lt;/sup&gt; I total</td>
<td>10.81 (6.65)</td>
<td>-.05 (.82)</td>
<td>11.11 (5.42)</td>
<td>.12 (.55)</td>
</tr>
<tr>
<td><strong>UPDRS II total</strong></td>
<td>13.14 (9.14)</td>
<td>-.04 (.87)</td>
<td>10.54 (7.39)</td>
<td>.35 (.07)</td>
</tr>
<tr>
<td><strong>UPDRS III total</strong></td>
<td>29.67 (17.46)</td>
<td>-.08 (.73)</td>
<td>28.85 (15.28)</td>
<td>.20 (.32)</td>
</tr>
<tr>
<td><strong>UPDRS IV total</strong></td>
<td>5.76 (3.90)</td>
<td>.23 (.31)</td>
<td>5.93 (4.60)</td>
<td>.07 (.71)</td>
</tr>
<tr>
<td><strong>UPDRS total</strong></td>
<td>59.38 (30.74)</td>
<td>.02 (.92)</td>
<td>56.70 (28.10)</td>
<td>.21 (.29)</td>
</tr>
<tr>
<td><strong>Zarit total</strong></td>
<td>14.67 (9.90)</td>
<td>.28 (.21)</td>
<td>8.92 (6.93)</td>
<td>.21 (.31)</td>
</tr>
<tr>
<td><strong>EQ-5D-5L item 4 (pain/discomfort)</strong></td>
<td>1.81 (0.87)</td>
<td>-.07 (.76)</td>
<td>2.64 (0.95)</td>
<td>-.11 (.58)</td>
</tr>
<tr>
<td><strong>NMSS item 11 (flat moods)</strong></td>
<td>1.33 (1.96)</td>
<td>.11 (.65)</td>
<td>1.07 (2.22)</td>
<td>.18 (.37)</td>
</tr>
<tr>
<td><strong>NMSS item 26 (problems having sex)</strong></td>
<td>2.24 (4.19)</td>
<td>-.31 (.17)</td>
<td>0.82 (3.14)</td>
<td>.12 (.57)</td>
</tr>
<tr>
<td><strong>PDQ-8 item 7 (painful cramps or spasms)</strong></td>
<td>1.70 (1.22)</td>
<td>-.005 (.99)</td>
<td>2.59 (1.48)</td>
<td>.03 (.90)</td>
</tr>
<tr>
<td><strong>UPDRS item 21 (speech)</strong></td>
<td>1.43 (1.03)</td>
<td>-.09 (.69)</td>
<td>0.75 (0.93)</td>
<td>.14 (.48)</td>
</tr>
<tr>
<td><strong>UPDRS item 33a (rigidity)</strong></td>
<td>1.19 (0.98)</td>
<td>.36 (.11)</td>
<td>0.54 (0.64)</td>
<td>.08 (.71)</td>
</tr>
<tr>
<td><strong>Zarit item 8 (social life suffered)</strong></td>
<td>1.29 (1.19)</td>
<td>.24 (.30)</td>
<td>0.62 (0.80)</td>
<td>.16 (.44)</td>
</tr>
</tbody>
</table>

<sup>a</sup>MMSE: Mini-Mental State Examination.
<sup>b</sup>EQ-5D-5L: EuroQol 5-Dimension 5-Level.
<sup>c</sup>NMSS: Nonmotor Symptoms Scale.
<sup>d</sup>PDQ-8: Parkinson’s Disease Questionnaire–8.
<sup>e</sup>UPDRS: Unified Parkinson’s Disease Rating Scale.

### Determinants of Compliance

A Kruskal-Wallis H test was run to determine if there were differences in use between patients’ groups (low, moderate, and high use) based on demographics and total scores as well as on their scoring in the ordinal variables which are indicating symptoms (NMSS and UPDRS items), quality of life aspects (PDQ-8 and EQ-5D-5L items) and caregiver burden reasons (Zarit items). The distributions of use were significantly different between groups for specific items of the scales and not for the total scores (Table 3).

A linear regression was run to understand the effect of caregivers’ burden on system use. Linearity was assessed with a scatterplot of Zarit_total against system use in which the regression line was plotted. Visual inspection of these two plots indicated a linear relationship between the variables. There was homoscedasticity and normality of the residuals. There were no outliers.

https://mhealth.jmir.org/2020/6/e16414

---

Gatsios et al. JMIR MHEALTH AND UHEALTH 2020 | vol. 8 | iss. 6 | e16414 | p.261
The prediction equation was: \( \text{use} = 48.31 + 1.51 \times \text{Zarit\_total} \). Zarit\_total statistically significantly predicted use, \( F_{1,59} = 5.86, P < .02 \), accounting for 30% of the variation in use with adjusted \( R^2 = 7.5\% \), a small size effect according to Cohen.

A multiple regression analysis was run to determine how much of the variation in the system use can be explained by the caregivers’ burden (Zarit total), Motor Aspects of Experiences of Daily Living (UPDRS II), and patients self-rated health status (EQ-5D-5L).

There was linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.855, and homoscedasticity, as assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There were no studentized deleted residuals greater than 3 standard deviations and values for the Cook distance above 0.2 (outliers). The assumption of normality was met, as assessed by a Q-Q plot. The multiple regression model statistically significantly predicted use, \( F_{3,56} = 5.650, P = .002 \). \( R^2 \) for the overall model was 48.2% with an adjusted \( R^2 \) of 19.1%, a medium size effect according to Cohen. All three variables added statistically significantly to the prediction, \( P < .05 \). Regression coefficients and standard errors can be found in Table 4.

A binomial logistic regression was performed to ascertain the effects of caregivers’ burden (Zarit total), Motor Aspects of Experiences of Daily Living (UPDRS II), and patients’ self-rated health status (EQ-5D-5L) on the likelihood of high system use. Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box-Tidwell procedure. A Bonferroni correction was applied using all 8 terms in the model resulting in statistical significance being accepted when \( P < .008 \). Based on this assessment, all continuous independent variables were found to be linearly related to the logit of the dependent variable. There was one standardized residual. The logistic regression model was statistically significant (\( \chi^2 = 13.5, P = .004 \)). The model explained 33.3% (Nagelkerke \( R^2 \)) of the variance in use and correctly classified 74.5% of cases. Sensitivity was 80.8%, specificity was 66.7%, positive predictive value was 75.0%, and negative predictive value was 73.3%. Of the three predictor variables, two were statistically significant: caregivers’ burden and patients’ self-rated health status as shown in Table 5. Users with better self-rated health status had 1.5 times higher odds to exhibit higher system use. Moreover, increasing caregivers’ burden was associated with an increased likelihood of higher system use.

### Table 4. Summary of multiple regression analysis.

<table>
<thead>
<tr>
<th>Model</th>
<th>B (^a)</th>
<th>SE(B) (^b)</th>
<th>Beta (^c)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>89.084</td>
<td>16.345</td>
<td>—</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>EQ-5D-5L (^d) total</td>
<td>-6.022</td>
<td>1.925</td>
<td>-0.465</td>
<td>.003</td>
</tr>
<tr>
<td>Zarit total</td>
<td>1.651</td>
<td>0.687</td>
<td>0.331</td>
<td>.02</td>
</tr>
<tr>
<td>UPDRS II total</td>
<td>1.757</td>
<td>0.847</td>
<td>0.326</td>
<td>.04</td>
</tr>
</tbody>
</table>

\(^a\)B: unstandardized regression coefficient.
\(^b\)SE\(B\): SE of the coefficient.
\(^c\)Beta: standardized coefficient.
\(^d\)EQ-5D-5L: EuroQol 5-Dimension 5-Level.

### Table 5. Summary of binary logistic regression.

<table>
<thead>
<tr>
<th>Model</th>
<th>B (^a)</th>
<th>SE(B) (^b)</th>
<th>Wald (\chi^2)</th>
<th>(P) value</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ-5D-5L (^c) total</td>
<td>0.396</td>
<td>0.158</td>
<td>6.252</td>
<td>.01</td>
<td>1.485</td>
</tr>
<tr>
<td>UPDRS (^d) II total</td>
<td>-0.088</td>
<td>0.057</td>
<td>2.395</td>
<td>.12</td>
<td>0.916</td>
</tr>
<tr>
<td>Zarit total</td>
<td>-0.122</td>
<td>0.053</td>
<td>5.186</td>
<td>.02</td>
<td>0.885</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.294</td>
<td>1.119</td>
<td>1.338</td>
<td>.25</td>
<td>0.274</td>
</tr>
</tbody>
</table>

\(^a\)B: unstandardized regression coefficient.
\(^b\)SE\(B\): SE of the coefficient.
\(^c\)EQ-5D-5L: EuroQol 5-Dimension 5-Level.
\(^d\)UPDRS: Unified Parkinson’s Disease Rating Scale.
Clinically Meaningful Data With Ecological Validity

The method for the evaluation of tremor was presented in Rigas et al [19]. The limitation of this method was that the validation of accuracy was done with annotations by clinicians over specific, short periods in the controlled environment of a clinic, following a specific protocol [16]. In this first data collection study, the sessions were filmed in order to validate the annotations with external observers. With the data collected in the pilot study presented here, we were able to evaluate whether the method works for patients performing daily activities. Video at home was excluded due to study participants’ privacy concerns. The annotation was the perceived tremor as depicted in UPDRS item 2.10 (which indicates how the patient experienced tremor over the past week), the rest tremor amplitude in the left and right upper extremity as depicted in UPDRS item 3.17 (which allows the rater to gather observations on rest tremor that may appear at any time during the exam), and the constancy of rest tremor as depicted in UPDRS item 3.18 (which focuses on the constancy of rest tremor during the examination period when different body parts are variously at rest). All UPDRS items were assessed at the baseline visit (ie, before the pilot use of the system). Maximum of 3.17a and 3.17b referring to rest tremor amplitude in upper extremities was also estimated as part of the analysis.

A total of 50 cases were included in tremor analysis since for these cases more than 30 hours of sensor data were available from the pilot study and the results can be considered as reliable. The tremor was constantly evaluated at any moment data were available from the system. For the evaluation the method presented in Rigas et al [19] was used, and this is the score depicted in Table 6. Bivariate correlations between UPDRS items scored by the clinicians at baseline and the tremor score with our method were calculated with the Pearson product-moment correlation coefficient.

We notice that the mean score for no tremor is close to zero. This is due to the fact that some daily movements can simulate tremor and, as explained, the score was constantly calculated. Consistently, we noticed a small increase of mean score for slight tremor and a more significant increase for mild and moderate tremor.

A Welch t test was run to determine if there were differences in scores between groups and statistically significant differences confirm the discrimination between the no-tremor and tremor groups.

Moreover, there is a statistically significant, strong positive correlation between the tremor score and amplitude and constancy of tremor as evaluated at baseline by the clinicians and a moderate positive correlation with tremor as perceived by the patient.

Table 6. Correlations between Unified Parkinson’s Disease Rating Scale tremor-related items and our tremor method scores.

<table>
<thead>
<tr>
<th>UPDRS item</th>
<th>2.10 (tremor as perceived by the patient)</th>
<th>3.17-a (rest tremor amplitude—right upper extremity)</th>
<th>3.17-b (rest tremor amplitude—left upper extremity)</th>
<th>Max 3.17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases with UPDRS=0, n</td>
<td>21</td>
<td>39</td>
<td>38</td>
<td>32</td>
</tr>
<tr>
<td>Scores for UPDRS=0, mean (SD)</td>
<td>0.038 (0.034)</td>
<td>0.073 (0.156)</td>
<td>0.066 (0.116)</td>
<td>0.037 (0.032)</td>
</tr>
<tr>
<td>Cases with UPDRS=1, n</td>
<td>18</td>
<td>8</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Scores for UPDRS=1, mean (SD)</td>
<td>0.123 (0.226)</td>
<td>0.184 (0.248)</td>
<td>0.220 (0.329)</td>
<td>0.128 (0.172)</td>
</tr>
<tr>
<td>Cases with UPDRS&gt;1, n</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Scores for UPDRS&gt;1, mean (SD)</td>
<td>0.267 (0.330)</td>
<td>0.538 (0.383)</td>
<td>0.421 (0.383)</td>
<td>0.540 (0.377)</td>
</tr>
<tr>
<td>Area under the curve</td>
<td>0.643</td>
<td>0.887</td>
<td>0.783</td>
<td>0.871</td>
</tr>
<tr>
<td>Welch t test, P value for UPDRS=0 and UPDRS&gt;1</td>
<td>0.005</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Welch t test, P value for UPDRS=0 and UPDRS=1</td>
<td>0.108</td>
<td>0.117</td>
<td>0.03</td>
<td>0.008</td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>.378</td>
<td>.544</td>
<td>.468</td>
<td>.711</td>
</tr>
<tr>
<td>Pearson P value</td>
<td>.007</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*UPDRS: Unified Parkinson’s Disease Rating Scale.

Discussion

Principal Findings

The most important finding of our study is that patients with moderate PD, regardless of their age, gender, education, severity of symptoms, specific symptoms, perceived quality of life, or caregiver burden, were compliant with the use the system for 1 to 2 weeks. Overall, 87% (65/75) of study participants were data contributors for 4.53 hours on average per day. They used the system for a median 11.57 (SD 3.15) days.

Regarding the compliance determinants, regression analysis suggests that the best predictor associated with system use was caregiver burden. The higher the burden the higher the use, a finding emphasizing the role of caregivers in adherence to mHealth solutions including wearables. Moreover, deterioration of a caregiver’s social life seemed to be the most influential factor among Zarit items. The implication of these findings is that the moderate use group demonstrated the lower caregiver burden.
Motor Aspects of Experiences of Daily Living (UPDRS II) also affect the use of the system with users facing several motor problems in their activities of daily life belonging to the high use group and the rest decreasingly in the moderate and low groups. Speech problems especially seem to discriminate groups. Moreover, patients self-rated health status seems to predict high and moderate use. Feeling pain or discomfort was the strongest individual predictor.

Another objective of the study was to collect data that are clinically meaningful (ie, data that the clinicians can use for monitoring and evaluation of symptoms when the patient is in his or her home environment. In this study, we provide evidence of clinical validity and ecological effect of an algorithm derived from a single sensor on the wrist for detecting tremor in PD patients. The applications of such monitoring methods include patients who cannot properly report their symptoms either because they are newly diagnosed or because they find it difficult to characterize tremor or even differentiate tremor from dyskinesias.

Following the paradigm of recent studies, PD_manager has built a large database for future development and testing of novel algorithms applied to sensor-derived data from PD patients during daily functioning. In total, more than 2700 hours of useful sensor data from the smartphone and Microsoft Band were collected and can be used for evaluating gait, freezing of gait, bradykinesia, tremor and dyskinesia, or monitoring and evaluation of fluctuations in future studies.

**Comparison With Prior Work**

This study contributes to the growing evidence about the feasibility of mHealth for PD patients. It is aligned with the findings that there are no noteworthy variances in baseline characteristics (age, gender, education, disease duration, and severity) that can explain compliance even in larger studies [4]. Findings such as highest compliance of older participants in one study [9], which can be attributed to more severe disease status and increased need for better management, and a negative impact of patients’ and caregivers’ education in this study, which can be the result of the lack of direct feedback from the system leading to limited self-management value, are worthy of further exploration.

By including patients with moderate disease severity (H&Y of 3) and by exploring the determinants of their compliance, PD_manager complements most prior studies that recruited mostly patients mildly affected (H&Y of 2 or less) [4,9-11,13]. The high level of system use and compliance of these more affected patients, as in previous studies, can be linked to factors including the simple and passive design of the patient’s app, which was basically providing a series of reminders for short motor and nonmotor tasks, the insight in the condition that the patients and their caregivers expect as a result of using the system, and the fact that the technology is considered an extension of prescribing clinicians and thus very important for better care. PD_manager was used as a PD Holter (ie, in a similar context as sensors used in recent studies [8,10,13]), and this complements the findings from previous research [3-5,9] that suggests that mHealth systems could be used both for short (1-2 weeks) and long-term (6 months) monitoring of PD patients.

Moreover, our findings are consistent with the Movement Disorder Society Task Force on Technology roadmap [20] as well as with patient attitudes on technology use [21]. Our mHealth platform, as relevant studies suggest, can be an effective tool for the passive, unobtrusive monitoring and evaluation of symptoms [22], defining new phenotypical biomarkers [23], detection of serious events such as falls [24], detection of worsening in the overall health status of the patients, and the provision of better disease management and improved care [25], the latter being already extensively studied in ongoing clinical trials (eg, NCT03741920 and NCT02657655). mHealth may also help rehabilitation [26,27] and facilitate telemedicine since it enables home-based [28], multidisciplinary [29] approaches for the management of PD. Moreover, the system could be used for connecting and sharing health data promoting research in PD [30] in line with EU priorities for enabling the digital transformation of health and care. Empowering citizens and promoting self-management is another important benefit of mHealth for PD patients [5]. Finally, mHealth can be used to provide decision support on the need for advanced treatments and their titration when they are applied [31].

**Limitations**

Limitations include the number of patients that used the system which, despite the excellent stratification that was preferred in this study over extended recruitment, should be increased in future studies to further establish the findings. The relatively preserved cognitive condition of study patients could be considered a limitation since cognitive deficits are common in advanced PD. Compliance was not calculated as the median percentage of the study period where accelerometer data were collected as in previous studies because the designs are different, and the technology limitations imposed a rather personalized use of the system during the waking day. Another limitation is that compliance should also be assessed in repeated 10- to 14-day periods at least twice a year as clinically meaningful use would demand to evaluate the long-term effects in patients’ care. Finally, more workshops with clinicians for improving the use of the system in clinical practice are required.

**Conclusions**

mHealth for monitoring of PD patients’ symptoms is feasible, at least for a period of 2 weeks. With the data collected with mHealth, ecologically valid, accurate, and objective monitoring and evaluation of tremor and other symptoms is feasible, and future studies should confirm its efficiency to support clinical decisions and improve patients’ management. Future mHealth systems should take into consideration and address the determinants of mHealth use, which include the subjective caregiver burden and especially the impact on social life, the self-evaluation of the activities of daily life including speech, and the overall patients’ self-rated health status with emphasis in pain and discomfort.
References


Abbreviations

EQ-5D-5L: EuroQol 5-Dimension 5-Level
H&Y: Hoehn and Yahr scale
NMSS: Non-motor symptoms scale
PD: Parkinson disease
PDQ-8: Parkinson's Disease Questionnaire-8
UPDRS: Unified Parkinson's Disease Rating Scale
Feasibility and Utility of mHealth for the Remote Monitoring of Parkinson Disease: Ancillary Study of the PD_manager Randomized Controlled Trial


Please cite as:
JMIR Mhealth Uhealth 2020;8(6):e16414
URL: https://mhealth.jmir.org/2020/6/e16414
doi: 10.2196/16414
PMID: 32442154

©Dimitris Gatsios, Angelo Antonini, Giovanni Gentile, Andrea Marcante, Clelia Pellicano, Lucia Macchiusi, Francesca Assogna, Gianfranco Spalletta, Heather Gage, Morro Touray, Lada Timotijevic, Charo Hodgkins, Maria Chondrogiorgi, George Rigas, Dimitrios I Fotiadis, Spyridon Konitsiotis. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 29.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Web-Based Mobile App (INTERACCT App) for Adolescents Undergoing Cancer and Hematopoietic Stem Cell Transplantation Aftercare to Improve the Quality of Medical Information for Clinicians: Observational Study

Anita Lawitschka, MD, PhD; Stephanie Buehrer, MSc; Dorothea Bauer, MD; Konrad Peters, MSc; Marisa Silbernagl, MSc; Natalia Zubarovskaya, MD; Barbara Brunnmaier, PhD; Fares Kayali, Prof Dr; Helmut Hlavacs, UnivProf; Natalia Zubarovskaya, MD; David Riedl, PhD; Gerhard Rumpold, Prof Dr; Christina Peters, MD; Barbara Brunmair, PhD; Fares Kayali, Prof Dr; Helmut Hlavacs, UnivProf; David Riedl, PhD; Gerhard Rumpold, Prof Dr; Christina Peters, MD

Corresponding Author:
Anita Lawitschka, MD, PhD
Stem Cell Transplantation-Outpatient and Aftercare Clinic
St. Anna Children’s Hospital
Medical University Vienna
Kinderspitalgasse 15
1090 Vienna
Austria
Phone: 43 1 40 170 ext 2900
Email: anita.lawitschka@stanna.at

Abstract

Background: A growing number of cancer and hematopoietic stem cell transplant (HSCT) survivors require long-term follow-up with optimal communication schemes, and patients’ compliance is crucial. Adolescents have various unmet needs. Regarding self-report of symptoms and health status, users of mobile apps showed enhanced compliance. Currently, HSCT aftercare at the HSCT outpatient clinic of the St. Anna Children’s Hospital in Vienna, Austria, is based on handwritten diaries, carrying various disadvantages. Recently, we developed the prototype of a web-based, self-monitoring gamified mobile app tailored for adolescents: the INTERACCT (Integrating Entertainment and Reaction Assessment into Child Cancer Therapy) app.

Objective: This observational, prospective study evaluated the usability of the INTERACCT app for tracking real-time self-reported symptoms and health status data in adolescent HSCT patients and a healthy matched control group. The primary outcome of the study was the quality of the self-reported medical information. We hypothesized that the mobile app would provide superior medical information for the clinicians than would the handwritten diaries.

Methods: Health data were reported via paper diary and mobile app for 5 consecutive days each. The quality of medical information was rated on a 5-point scale independently and blinded by two HSCT clinicians, and the duration of use was evaluated. A total of 52 participant questionnaires were assessed for gaming patterns and device preferences, self-efficacy, users’ satisfaction,
acceptability, and suggestions for improvement of the mobile app. Interrater reliability was calculated with the intraclass correlation coefficient, based on a two-way mixed model; one-way repeated-measures analysis of variance and t tests were conducted post hoc. Descriptive methods were used for correlation with participants’ demographics. For users’ satisfaction and acceptability of the mobile app, the median and the IQR were calculated.

Results: Data from 42 participants—15 patients and 27 healthy students—with comparable demographics were evaluated. The results of our study indicated a superiority of the quality of self-reported medical data in the INTERACCT app over traditional paper-and-pencil assessment (mobile app: 4.14 points, vs paper-based diary: 3.77 points, P=.02). The mobile app outperformed paper-and-pencil assessments mainly among the patients, in particular among patients with treatment-associated complications (mobile app: 4.43 points, vs paper-based diary: 3.73 points, P=.01). The mobile app was used significantly longer by adolescents (≥14 years: 4.57 days, vs ≤13 years: 3.14 days, P=.03) and females (4.76 days for females vs 2.95 days for males, P=.004). This corresponds with a longer duration of use among impaired patients with comorbidities. User satisfaction and acceptability ratings for the mobile app were high across all groups, but adherence to entering a large amount of data decreased over time. Based on our results, we developed a case vignette of the target group.

Conclusions: Our study was the first to show that the quality of patient-reported medical information submitted via the INTERACCT app embedded in a serious game is superior to that submitted via a handwritten diary. In light of these results, a refinement of the mobile app supported by a machine learning approach is planned within an international research project.

(JMIR Mhealth Uhealth 2020;8(6):e18781) doi:10.2196/18781

KEYWORDS
mobile app; adolescents; cancer; stem cell transplant; self-reported health status; medical information exchange; mobile phone

Introduction

Allogeneic hematopoietic stem cell transplantation (HSCT) is a curative treatment for an increasing range of mostly malignant hematopo-oncological diseases. Due to progress in transplant procedures, a growing number of survivors are subject to many comorbidities and late effects, which are associated both to treatment and related complications [1,2]. Among the latter, acute and chronic graft-versus-host disease (cGVHD) are the most significant, affecting survival and multidimensional aspects of quality of life [3-5]. Survivors require long-term follow-up care by highly specialized clinics, with tight communication and clinical control schemes [6]. In turn, patients’ motivation and compliance are crucial factors for the success of aftercare and rehabilitation.

It is now widely accepted that teenage and young adult patients undergoing cancer treatments and HSCT have a number of informational and psychosocial needs that are frequently not addressed. These needs are practical as well as being related to counseling and coping strategies [7]. Moreover, loss of compliance has been shown among adolescent cGVHD survivors during long-term follow-up, indicating that there is an unmet need to enhance monitoring of comorbidities, graft-versus-host disease (GVHD) symptoms, and quality of life during aftercare in adolescents [8].

Currently, adolescent HSCT aftercare at the St. Anna Children’s Hospital in Vienna is based on a traditional paper diary, which carries important disadvantages that are in line with reports by others: (1) the adherence by patients is low, (2) the diaries are often not filled regularly, (3) the accuracy of symptom recording diminishes when recall periods extend (ie, recall bias), and (4) health data are entered retrospectively and only prior to the clinic visit (ie, fake compliance) [9,10]. These disadvantages may be prevented by real-time electronic symptom report (eg, via smartphone or tablet apps adding tracking abilities) [9-11].

Adolescents are frequent technology consumers and feel more comfortable with the electronic exchange of sensitive information [10]. Today, many opportunities exist to record health-related data due to the development of new technologies [12]. Real-time symptom recording may provide the possibility to detect negative health changes earlier, independent of hospital visits, providing the possibility to initiate coping strategies [13]. In summary, the use of mobile apps for symptom self-report satisfies patients’ preferences, improves compliance, and reduces errors in data collection [9,14-17]. It is well accepted that gamification approaches may enhance users’ motivation and satisfaction, along with compliance and ability for self-management [16,18].

Most mobile app publications for adolescent cancer patients have focused on pain; less data are available for HSCT patients. Generally, the majority of studies examined usability, acceptability, and adherence to app use. Although there is now a plethora of health mobile apps available [19], a review published in 2015 revealed only four pilot-tested mobile apps for young cancer patients [20]. Medical professionals are rarely included in the development of health-related software [13]. A review by Payne et al outlined the need to determine the efficacy of mobile apps [21]. To our knowledge, no study has compared the quality of medical information for the clinician when obtained via a mobile app with the information content obtained via a handwritten paper diary.

Recently, we have developed a prototype of a web-based, self-monitoring, gamified smartphone app, designed for adolescents in particular, as part of a third-party funded project—INTERACCT (Integrating Entertainment and Reaction Assessment into Child Cancer Therapy) [22]—together with academic partners from the University of Vienna, Entertainment Computing Research Group, and the University of Applied Arts Vienna, as well as the industry partner T-Systems Austria. To enhance motivation for data submission by users, the...
INTERACCT app was co-designed with adolescents and embedded into a serious game.

This study describes the evaluation of the INTERACCT app for tracking real-time self-reported symptoms and health status data in adolescents after HSCT in comparison with a healthy age- and gender-matched control group. As a first step on the way to test the INTERACCT app within a randomized controlled trial, we defined the high quality of medical information transfer as an indicator of efficacy. Therefore, we wanted to determine whether the quality of medical information reported via the mobile app is at least as informative for the clinician as that reported by the traditional paper diary, which has been used so far. Within this observational study, funded by a grant from the Austrian Society for Hematology and Oncology, clinicians evaluated the quality of the reported medical information and compared it with the information obtained from handwritten diaries, both from patients and from an age-matched healthy control group. We hypothesized that the mobile app provides significantly superior medical information for the clinicians than the handwritten diaries.

**Methods**

**Participants**

The observational, prospective INTERACCT study was conducted among HSCT survivors during regular visits at the HSCT outpatient clinic of the St. Anna Children’s Hospital. Inclusion criteria were as followed: life expectancy of more than 6 months, at least 12 years of age, and no evidence of recurrence of primary disease. Healthy students from two classes at a Viennese school were recruited on a voluntary basis as the control group. Study briefing and instructions for use were given by an HSCT clinician visiting their school. Written informed consent was taken home, completed, and collected by the involved teacher. The study was conducted in accordance with the Declaration of Helsinki and has been approved by the institutional review board of the Medical University of Vienna and the St. Anna Children’s Hospital. Furthermore, the project was reviewed and approved by Vienna’s Board of Education (Stadtschulrat Wien, in German). All participants and parents were informed that the individual results of the project evaluations could not be disclosed. Written informed consent was obtained from participants and parents; 52 participants—16 patients (31%) and 36 healthy control group students (69%)—between the ages of 12 and 19 years were pseudonymously enrolled. No cointerventions or training sessions were given to the participants.

Participants’ sociodemographic data were self-reported, and patients’ medical data were retrieved from medical records. Cancer treatment–associated (ie, for malignant underlying diseases) severe comorbidity was classified according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) v3.0 [23] with a grade of 3-4. The HSCT-associated complication cGVHD was classified according the National Institutes of Health consensus criteria for cGVHD [24].

Participants used the paper diary first for 5 consecutive days, from Monday to Friday, followed by 5 days using the mobile app, also from Monday to Friday, to report given health status parameters and symptoms. The study included a no-waiting-list design. An overview of this observational study is shown as a diagram adapted from the CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram (see Figure 1) [19,25].
Figure 1. Study overview. The diagram was adapted from the CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram (Eysenbach, 2011, and CONSORT, 2020). Malignant diseases include all malignant underlying diseases. Comorbidity is deemed severe when graded as 3-4 on the Common Terminology Criteria for Adverse Events (CTCAE), cGVHD: chronic graft-versus-host disease; HSCT: hematopoietic stem cell transplantation; Q: question; T1: time point 1 at study entry; T2: time point 2 at study termination.

**Instruments**

The primary outcome of the study was the quality of the self-reported medical information, which was rated independently and blinded by two experienced HSCT physicians on a 5-point scale, ranging from 1 (low medical information content) to 5 (high medical information content). The traditional paper diary asked for seven domains to be reported on: stool frequency and quality, urinary abnormalities, nausea and vomiting, body temperature, pain, well-being, and activity—pain and well-being were scored on a 3-point scale. No further specific questions have been raised, but users had the possibility to add comments freely. The medical information collected in
the mobile app included 12 domains. See Figure 2 for more details.

To further assess the usability of the measure, several aspects were investigated. For assessment of user satisfaction, an adapted version of the Game Experience Questionnaire was employed [26]; items were measured on a 3-point Likert scale, ranging from 1 (not at all) to 3 (extremely). Participants were asked to report how bored, impressed, frustrated, tired, irritated, skillful, satisfied, challenged, strained, and well they felt during the mobile app usage. App-design acceptability was assessed by participants on a 3-point scale, ranging from 1 (worst) to 3 (best), for each of the 12 health domains and symptoms (see Figure 2). Additionally, participants were asked to share suggestions for improvements of the mobile app.

Figure 2. INTERACCT (Integrating Entertainment and Reaction Assessment into Child Cancer Therapy) app screenshots. A: Icons for the domains of self-reported health status—fluid and food intake (amount and difficulties); urine and stool (frequency, quality, and pain); physical exercise and play (duration in minutes); nausea and vomiting (frequency); fatigue (frequency); mouth complaints (frequency, quality, and pain); body temperature; and skin complaints (localization, frequency, and quality). B: Input for amount of fluid intake and difficulties, as an example.

For capturing patterns of gaming behavior, we designed an in-house questionnaire. Participants were asked to rate how often they play video games alone or with friends and how often they use certain devices on a 3-point Likert scale, giving a score of 1 (less than once a week), 2 (about once a week), or 3 (several times a week or daily). Finally, the average time spent, daily, on game activities was requested.

Self-efficacy of participants was assessed using a modified version of the Jerusalem Scales, in order to uncover the psychosocial background of participants. Self-efficacy questionnaires were requested at study entry (T1) and termination (T2) as outlined in the study overview [27]. The self-efficacy scales are widely used and target self-beliefs, cognitive and behavioral strategies necessary for problem solving (ie, coping), positive perception of the future (ie, optimism), as well as individuals’ appraisal of stress (ie, challenge, threat, and loss perception); items are scored on a 4-point Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree) [27].

Technical Components

The technical components were comprised of a Unity 3D 8 app (Unity Technologies) for the participants and an ASP.NET 9 web interface (Microsoft) for the clinicians. The persistence layer was implemented through the ASP.NET back end, accessing a Microsoft SQL (Structured Query Language) database, running in a secure Windows Server 2016 environment. The connection between clients and server was established using Secure Sockets Layer (SSL). Basic authentication and cross-site request forgery tokens were used to ensure data integrity and security. The INTERACCT client is a native smartphone app for Android and iOS. The main components of the client are remote medical data entry (ie, self-reported symptom and health status) as well as the game content itself. Scoring of symptoms mostly followed a 0-3 scale.
and was accumulated by the back end for each day. Figure 2 shows details of the mobile app with the health status domains (A) and, as an example, the input per amount of fluid intake and difficulties, thereby (B). When completing the data entry, users were rewarded with virtual in-game currency, used to progress in the main game. The core game idea of INTERACCT is to collect avatars and complete procedurally generated levels. The levels come in different graphical settings with different characters, in order to provide variety. A detailed developmental project description, as well as design considerations and preliminary results, are described by Kayali et al, Peters et al, and Mateus-Berr et al [28-31].

**Statistical Analysis**

The aim of the study was to determine whether the quality of electronically collected, self-reported medical information was comparable to that of traditional paper-and-pencil self-assessment. The participants first used the paper-and-pencil self-report form for 5 consecutive days, followed by a 5-day period using the INTERACCT app. The physician-rated quality of the self-reported medical information in both scenarios was used as primary outcome measure. To investigate differences in the primary outcome, one-way repeated-measures analyses of variance (ANOVAs) were calculated. Only patients with complete datasets were included in the analyses. The between-subject factor group was included to examine any differences between patients and students. We used *t* tests for associated samples as post hoc tests. Since, to our best knowledge, no prior studies have compared the quality of self-assessed medical information between paper-and-pencil and electronic assessments, and due to the observational nature of this pilot study, no a priori sample size calculations were feasible. Yet, to investigate whether the main results of the study are valid, a post hoc power analysis was conducted using G*Power 3.1 (Heinrich-Heine-Universität). To evaluate the interrater reliability of the physician-based evaluations, intraclass correlation coefficients (ICCs), based on a two-way mixed model and the absolute agreement, were calculated. The primary outcome measure was correlated with participants’ demographics—age ≤13 years versus ≥14 years, gender, and native language German or other—to investigate potential confounding effects. Due to the small sample sizes of these groups, descriptive methods (ie, mean and SD) were preferred over interference statistics. For descriptive correlation with patients’ demographics, subgroups were defined as follows: malignant versus nonmalignant underlying diseases, cGVHD (yes vs no), and severe comorbidity (CTCAE grade of 3-4) versus no comorbidity or mild comorbidity (CTCAE grade of 0-2) [23]. For assessment of users’ satisfaction and acceptability of the mobile app design, the median and IQR were calculated for each item.

**Results**

**Participants**

Only participants who used both the paper diary and the mobile app for 2 or more days and completed all questionnaires were included in the analyses. Data from 1 patient (due to death) and 9 students (due to incomplete data) had to be excluded, leaving data from 42 participants for further analyses: 15 patients, mainly after HSCT for malignant underlying diseases, and 27 healthy students (see Figure 1). Table 1 shows that the participants’ characteristics and demographic details—age, gender, and native language—were comparable (data not shown). Note that mean age was higher (data not shown) for some patient subgroups (ie, patients with malignant underlying diseases, cGVHD, and severe comorbidities).
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients (N=15)</strong></td>
<td></td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>8 (53)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>14.40 (2.29)</td>
</tr>
<tr>
<td>Age (≥14 years), n (%)</td>
<td>9 (60)</td>
</tr>
<tr>
<td>Native language (German), n (%)</td>
<td>12 (80)</td>
</tr>
<tr>
<td><strong>Malignant underlying disease (n=10)</strong></td>
<td></td>
</tr>
<tr>
<td>Total, out of all 15 patients, n (%)</td>
<td>10 (67)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>15.20 (3.97)</td>
</tr>
<tr>
<td>Time since HSCT(^a) in years, mean (SD)</td>
<td>3.22 (4.53)</td>
</tr>
<tr>
<td><strong>cGVHD(^b) (n=5)</strong></td>
<td></td>
</tr>
<tr>
<td>Total, out of all 15 patients, n (%)</td>
<td>5 (33)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>5 (100)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>16.20 (2.78)</td>
</tr>
<tr>
<td><strong>Comorbidity: CTCAE(^c) grade 3-4 (n=8)</strong></td>
<td></td>
</tr>
<tr>
<td>Total, out of all 15 patients, n (%)</td>
<td>8 (53)</td>
</tr>
<tr>
<td>Grade of 3, n (%)</td>
<td>5 (63)</td>
</tr>
<tr>
<td>Grade of 4, n (%)</td>
<td>3 (37)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>5 (63)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>16.00 (2.73)</td>
</tr>
<tr>
<td><strong>Healthy control group (N=27)</strong></td>
<td></td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>13 (48)</td>
</tr>
<tr>
<td>Age in years, mean (SD)</td>
<td>13.22 (1.12)</td>
</tr>
<tr>
<td>Native language (German), n (%)</td>
<td>19 (70)</td>
</tr>
</tbody>
</table>

\(^a\)HSCT: hematopoietic stem cell transplantation.  
\(^b\)cGVHD: chronic graft-versus-host disease.  
\(^c\)Classification according to the National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) [23].

Analysis of Patterns of Gaming Behavior

Our results showed different patterns of gaming behavior between patients and the healthy control group. Generally, the playing frequency was found to be higher in patients than in students, with patients playing daily or at least several times a week (10/15, 67% vs 12/27, 44%, respectively). This finding was more pronounced in males, in older participants (ie, ≥14 years of age), and in participants whose native language is German.

The majority (10/15, 67%) of patients reported playing computer games alone rather than with friends (1/15, 7%). This applied particularly to the more-impaired patient groups with cGVHD and comorbidities (CTCAE grade of 3-4), of which not a single person indicated playing games with friends. In contrast, 10 out of 27 (37%) students from the control group reported playing with friends. Most of the participants spent an average of 2 hours (up to 2.65 hours) per day playing computer games. Notably, even the younger users (ie, ≤13 years of age) reported using mainly the smartphone for about 2 hours a day. The most preferred kind of personal computing device was clearly the smartphone among the majority of the users (patients: 10/15, 67%, and healthy controls: 16/27, 59%); this was mostly evident in the patient group with malignant diseases, cGVHD, and severe comorbidities (see Multimedia Appendix 1).

Comparison of the Paper Diary With the Mobile App

Overview

Over the entire study cohort, the clinicians rated the medical information content of the mobile app with an average of 4.14 (SD 1.14) points versus 3.77 (SD 0.91) points for the paper diary (see Figure 3). The repeated-measures ANOVA revealed a significant difference between the information content of the mobile app and the paper diary, with a medium-to-large effect size ($F_{1,40}=5.571, P=.02, \eta^2=.12$). Overall, results indicate that the mobile app provided significantly more information for the clinicians than the handwritten diary. Additionally, the...
between-subject factor group was not significant ($F_{1,40}=0.522, P=.47$) and had no significant interaction effect ($F_{1,40}=1.807, P=.19$). A post hoc power analysis revealed a power of $\beta=0.99$ for the repeated-measure ANOVA, indicating a sufficiently powered sample for this analysis.

**Figure 3.** Comparison of the paper diary with the mobile app. The plot shows the clinician-reviewed quality of medical information regarding patient-reported health status on a 5-point scale, ranging from 1 (low) to 5 (high). Malignant refers to malignant underlying disease. cGVHD: chronic graft-versus-host disease; CTCAE 3-4: National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) grade of 3-4.

Post hoc t tests confirmed that the significant main effect was predominantly driven by the patients. The patients received significantly higher average points for the mobile app than for the paper diary (mobile app: mean 4.43, SD 1.07, vs paper diary: mean 3.73, SD 0.63, $t_{1,14}=-2.941, P=.01$). No significant difference between the mobile app and the paper diary was found for the control group (mobile app: mean 3.98, SD 1.17, vs paper diary: mean 3.78, SD 1.05, $t_{1,26}=-0.772, P=.45$). No correlation with user-specific characteristics, such as age, gender, and native language, was observed. A tendency toward higher information content ratings via the mobile app in patients with malignant diseases was observed (see Figure 3). However, the mobile app always delivered higher ratings for the quality of medical information than did the handwritten diary.

**Quality of Reported Medical Information Content**

The quality of the self-reported medical information was rated independently by two blinded experienced HSCT clinicians on a 5-point scale, ranging from 1 (low medical information content) to 5 (high medical information content), and revealed a high interrater reliability among them (see Figure 3). The average ICC was .867 (95% CI .733-.932, $F_{41,41}=8.530, P<.001$) for the paper diary and .912 (95% CI .794-.958, $F_{41,41}=14.00, P<.001$) for the mobile app.

**Duration of Data Entry in Days**

As seen in Figure 4, patients used the paper diary more than required for an average of 5.67 (SD 1.23) days. The mobile app was used by this group for an average of 4.67 (SD 2.53) days, but this difference was not significant ($t_{1,14}=1.479, P=.16$). In contrast, the control group used the diary as long as the patients did, but the mobile app was used for a significantly shorter time than was the paper diary (mean 5.89 days, SD 1.31, vs mean 3.41 days, SD 1.69, respectively; $t_{1,26}=6.138, P<.001$).

Correlation analysis with users’ demographics showed that the mobile app, but not the paper diary, was used significantly longer by the adolescent group (≥14 years: 4.57 days, vs ≤13 years: 3.14 days, $t_{1,40}=-2.331, P=.03$) and females (4.76 days for the females vs 2.95 days for the males, $t_{1,40}=3.082, P=.004$).

The descriptive correlation with patient subgroup characteristics showed a positive correlation (ie, longer duration of data entry) with cGVHD (cGVHD: 6.00 days, vs no cGVHD: 4.00 days) and comorbidity (CTCAE 3-4: 5.00 days, vs CTCAE 0-2: 4.10 days).

**General User Satisfaction and Acceptability of the Mobile App Design**

User satisfaction for the mobile app was high across all groups. Table 2 shows the median score and the IQR for all the user satisfaction items.
Figure 4. Comparison of the paper diary versus mobile app: duration in days of data entry. Malignant refers to malignant underlying disease. cGVHD: chronic graft-versus-host disease; CTCAE 3-4: National Cancer Institute (NCI) Common Terminology Criteria for Adverse Events (CTCAE) grade of 3-4.

Table 2. User satisfaction of the mobile app.

<table>
<thead>
<tr>
<th>User satisfaction item</th>
<th>Score(^a), median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I was bored</td>
<td>2.00 (2.00)</td>
</tr>
<tr>
<td>I was impressed</td>
<td>2.00 (2.00)</td>
</tr>
<tr>
<td>I felt frustrated</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>I found it tiring</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>I was irritated</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>I felt skillful</td>
<td>2.00 (2.00)</td>
</tr>
<tr>
<td>I was satisfied</td>
<td>3.00 (3.00)</td>
</tr>
<tr>
<td>I felt challenged</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>I had to put a lot of effort into playing</td>
<td>1.00 (1.00)</td>
</tr>
<tr>
<td>I felt good</td>
<td>3.00 (3.00)</td>
</tr>
</tbody>
</table>

\(^a\)Scores range from 1 (not at all) to 3 (extremely).

In particular, we found that most participants (minimum 25/42, 60%, maximum 40/42, 95%) felt hardly frustrated, tired, irritated, or strained. On average, more than half of the participants stated that they were very satisfied during the mobile app usage, were hardly challenged, and felt good (see Multimedia Appendix 2).

Patient subgroups with malignant diseases, cGVHD, and CTCAE grade 3-4 representing the most experienced patients indicated that they were a little bored with the mobile app (see Multimedia Appendix 3). This finding corresponds with the result that relatively more patients (13/15, 87%) felt hardly challenged by the mobile app compared to the healthy control group (15/27, 56%). However, nearly half of the patients (7/15, 46%) were impressed.

Additionally, the acceptability of the app design on a 3-point Likert scale—1 (negative rating), 2 (neutral rating), or 3 (positive rating)—for each of the 12 domains revealed a median score of 3 (IQR 3). In summary, the app rating was encouraging, and there was no evidence that the diary was preferred to the mobile app.

All users were asked for suggestions for intervention and app improvement. Features desired by the users included (1) the insertion of comments and functions, such as scrolling within the app, (2) the overview of the entered data, making visible...
the progress or decline they have experienced, (3) the development of a more colorful design, (4) the implementation of captions for the symbols, (5) the integration of the fitness domains into the physical exercise domain, aiming for reduced parameters that they are asked about, and (6) a broader variety of health games.

**Self-Efficacy**

No significant differences in self-efficacy between user cohorts and time points (T1 vs T2) were observed using the Welch $t$ test. Descriptive comparison of patient subgroups revealed that the most impaired patients, with cGVHD and comorbidity CTCAE grade 3-4, reported lower scores for optimism, coping, and general self-efficacy, while reporting higher scores for loss and threat perception, in comparison to other patients. All findings remained stable over the period of the study (data not shown).

We summarized several findings of our study, which may be specific for the target group of impaired adolescent patients after cancer and HSCT, and created a case vignette for further use by developers and researchers (see Multimedia Appendix 4).

**Discussion**

**Principal Findings**

Recently, we have developed a web-based, self-monitoring mobile app for adolescent patients after cancer treatment and HSCT, within the multidisciplinary research project INTERACCT, at the interface of clinical research, design, and information technology [28-30]. This study describes the multidimensional testing of the co-designed INTERACCT app for tracking real-time self-reported symptoms and health status data in patients compared with a healthy age- and gender-matched control group. Our results show that mobile app use, when compared to the traditional paper diary, resulted in significantly improved quality of medical information for the clinicians. The mobile app outperformed the paper-and-pen assessment, particularly within the patient cohort; this was accentuated in more-impaired patients with treatment-associated complications, suggesting that the app is suitable for the group it was developed for. In accordance with the highly rated usability and acceptability, our results indicate that adolescents, after cancer treatment and HSCT, would use a real-time, self-monitoring smartphone app for tracking symptoms and changes in their health status, resulting in improved patient-physician communication.

As data submission through the INTERACCT app was embedded into a serious game, we also analyzed patterns of gaming habits and preferences for electronic devices; clearly, the smartphone is the appropriate device for our possibly chronically ill target group. Not surprisingly, the identified patterns of gaming habits may reflect the living situations of HSCT patients, who are often not participating in age-appropriate activities with peers. Our data suggest that, especially for adolescent patients with treatment-associated complications who are living a more isolated life, the implementation of a smartphone-based symptom-reporting tool seems appropriate, as described previously for cancer patients [18-20].

In addition to published app studies in cancer patients, which mostly described the development, implementation, and usability testing, we hypothesized that the quality of the medical information, which was submitted through a serious game, would be superior to the medical information collected via handwritten diary [32,33]. Our results confirm this hypothesis significantly for the patient group (mobile app: 4.43 points, vs paper-based diary: 3.73 points, $P=0.1$) and the finding is even more significant in cancer patients. The superiority of the mobile app is evident regarding the duration of symptom recording within the group of impaired adolescents after HSCT: symptom report via mobile app was significantly longer in patients over 13 years of age, female, with cGVHD, and with comorbidities with CTCAE grade 3-4 (cGVHD: 6.00 days, vs no cGVHD: 4.00 days; CTCAE 3-4: 5.00 days, vs CTCAE 0-2: 4.10 days). This may be due to several reasons: (1) symptom assessment 24/7 makes recording possible when the health problem arises, overcoming the problem of the recall period [9], and (2) motivation may have been enhanced by the serious game [18]. Fortier et al developed Pain Buddy for young cancer patients, providing an animated avatar, gamification components, and remote symptom monitoring, revealing high user satisfaction (n=12) while keeping track of their pain [34]. Baggott et al first published the results of utility testing of an electronic diary among 10 adolescent cancer patients, which not only reported pain but additionally reported nausea, vomiting, fatigue, sleep, and mood, and showed high adherence and reported benefit [10].

One may speculate that patients with more serious symptoms show better adherence comparable to diabetes patients [35]. Accordingly, the patients with malignant diseases, comprised of experienced patients without suffering from severe complications, scored better regarding the quality of medical information, even though they used the app for a shorter duration. Similarly, Rodgers et al reported an extensive decrease of app usage over time—the app EAT! assists in self-management of eating difficulties—in 16 HSCT patients [36]. Regarding the lower results of the mobile app usage within the control group, we speculate that the lack of symptoms hampers the conscientious documentation of health parameters; finally, the mobile app does not allow retrospective filling. Indeed, when we went back to the paper diaries of the control group, we found evidence for copy and paste in about 50% of paper diaries, comparable to the results by Stone et al (ie, fake compliance) [9].

In line with many published data, the usability and acceptance of the mobile app and its design was very positive and there was no evidence that the diary was preferred to the mobile app [18,20,37]. As indicated by Crosby et al, the participation of a multidisciplinary team combining experts of clinical research, design, and information technology may have been beneficial [38]. High compliance rates and acceptability in adolescent cancer patients have been reported by the use of Pain Squad as an iPhone app [16]. Subsequently, the group was able to show construct validity and reliability of the app [39] but, recently,
the implementation in a natural setting has outlined important challenges [40].

The descriptive results of the self-efficacy evaluation stress the importance of the implementation of a mobile app for improved patient-physician communication within the vulnerable HSCT patient group suffering from severe comorbidities and cGVHD.

Limitations
The main limitation of this study is related to the small sample size making statistical analyses of the subgroups impossible. Another drawback is that the intervention periods were quite short. Even though we used an age- and gender-matched control group, it seems very likely that the patients were more experienced regarding the self-report of symptoms. Additionally, the criteria for selecting the health domains with various symptoms for the INTERACCT app were partly empirically based and the number was perceived as too high—12 mobile app domains versus seven paper diary domains—by most users. Retrospectively, the time points of testing the self-efficacy seem to have been too close.

Although there is a myriad of reports about utility testing in mobile apps tracking various symptoms in cancer patients, Wesley and Fizur comprehensively outline the limitations, such as small sample numbers, poor correlation with sociodemographic parameters, and short intervention periods, and they highlight further research efforts [20].

Nevertheless, we have learned important lessons: patients after cancer treatment and HSCT represent a highly experienced target group regarding smartphone-based apps and health games. This has to be considered for further refinement as does the implementation of health games where the competition from thousands of available products is quite high. Pereira et al. reported that gamification elements enhance users’ motivation initially, but the interest declines over time [18]. The number of self-reported health domains requires careful consideration because of the wide variation in patients’ abilities to maintain attention to tasks. Hence, it might be useful to reduce the number of items in the app and to minimize the associated exertion. The high usability of the mobile app was not sufficient enough to foster sustained health data entry of our comorbid patients in the long term [41]. In addition to usability, illness experience, information technology infrastructure, emotional activation, degree of burden caused by the app, and relevance of symptom monitoring are important factors for patient motivation [42]. The less data that users are asked to enter, the greater their level of adherence in general [43]. Adaptive individual patient profiles, aiming for collection of relevant health data only, are warranted.

Implications of the Pilot Study
Our results have various implications: real-time assessment of symptom and health status changes may improve understanding of treatment-associated complications promoting enhanced supportive care. The model of real-time assessment described in this study can be used in both clinical and research contexts to evaluate the effectiveness of diagnosis and treatment. Ideally, it would move the field of clinical HSCT-aftercare research forward with identifying prodromes of complications in the future. INTERACCT already includes a web interface for clinicians to observe the submitted data, but its evaluation was not in the scope of this study. Plans are in place to test the refined mobile app within a multicenter randomized controlled study. We are planning not only to refine the INTERACCT app but, additionally, to apply machine learning algorithms aiming for implementation of adaptive individual patient profiles [44], thus enabling the mobile app to collect relevant health data and symptoms only.

Conclusions
In this study, we found support for its main hypothesis, that the quality of patient-reported medical information submitted via the INTERACCT app embedded in a serious game is superior to that from the traditional handwritten diary. The statistical results confirm this hypothesis. This is especially important for the target group of adolescent patients after cancer treatment and HSCT with severe comorbidities, where the results were even more conclusive than in the healthy control group. With this in mind, and given the scope for improvement in this relatively new research area, we plan to refine the use of the mobile app by considering machine learning approaches within an international multidisciplinary research project.

Acknowledgments
The authors would like to thank all study participants and supporting parents and teachers. The authors would like to acknowledge Thomas Pletschko, onco-psychologist at the Department of Pediatrics and Adolescent Medicine, Medical University Vienna, for his support regarding the self-efficacy questionnaire. This study was funded partly by the 2016 Occursus Prize of the OeGHO (Die Österreichische Gesellschaft für Hämatologie & Medizinische Onkologie), the Austrian Society of Hematology and Oncology.

Conflicts of Interest
None declared.

Multimedia Appendix 1

[File: mhealth_v8i6e18781_app1.png]
Multimedia Appendix 2
User satisfaction: percentage of users who felt very impressed, very skillful, very satisfied, hardly challenged, and very well.

Multimedia Appendix 3
User satisfaction: percentage of users who felt hardly bored, hardly frustrated, hardly tired, hardly irritated, and hardly strained.

Multimedia Appendix 4
Case vignette: specifics of impaired adolescent patients after cancer and hematopoietic stem cell transplantation (HSCT).

References


22. INTERACCT. URL: http://www.interact.at [accessed 2020-04-17]


Feasibility and Acceptability of a Mobile Mindfulness Meditation Intervention Among Women: Intervention Study

Ariane Lisann Rung1, PhD, MPH; Evrim Oral2, PhD; Lara Berghammer3, MPH; Edward S Peters1, DMD, SM, ScD

1Epidemiology Program, School of Public Health, Louisiana State University Health Sciences Center New Orleans, New Orleans, LA, United States
2Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center New Orleans, New Orleans, LA, United States
3Wounded Warrior Project, San Diego, CA, United States

Abstract

Background: Traditional mindfulness-based stress reduction programs are resource intensive for providers and time- and cost-intensive for participants, but the use of mobile technologies may be particularly convenient and cost-effective for populations that are busy, less affluent, or geographically distant from skilled providers. Women in southern Louisiana live in a vulnerable, disaster-prone region and are highly stressed, making a mobile program particularly suited to this population.

Objective: This study aimed to (1) assess the feasibility and acceptability of a mobile mindfulness app in real-world conditions in a pilot study of a community sample of women residing in southern Louisiana, (2) describe predictors of app usage, and (3) assess the effect of the app on secondary health outcomes.

Methods: Women were recruited from an oil spill study on health. A total of 236 women completed a baseline survey, were offered the mobile mindfulness program, and completed a follow-up survey. Subjects were asked to download and use the app for at least 30 days for 10 min. All study procedures were completed on the web. Primary outcomes were feasibility and acceptability of the app and characteristics of app utilization. Secondary outcomes included mindfulness, depression, perceived stress, sleep quality, physical activity, BMI, and healthy eating.

Results: Overall, 74.2% (236/318) of subjects completed the follow-up survey, and 13.5% (43/318) used the app. The main barrier to app usage was lack of time, cited by 37% (16/43) of users and 48.7% (94/193) of nonusers of the app. Women who chose to use the app were more highly educated (16/43, 63% had a college education vs 65/193, 33.7% of nonparticipants; P<.001), had higher incomes (23/43, 58% had incomes >US $50,000 per year vs 77/193, 43.0% of nonparticipants), and were employed (34/43, 79% vs 122/193, 63.2% of nonparticipants; P=.047). Those who engaged with the app did so at high levels, with 72% (31/43) of participants self-reporting the completion of some or all sessions and 74% (32/43) reporting high levels of satisfaction with the app. Participation with the app had a beneficial impact on depression (odds ratio [OR] 0.3, 95% CI 0.11-0.81), sleep quality (OR 0.1, 95% CI 0.02-0.96), sleep duration (OR 0.3, 95% CI 0.07-0.86), sleep latency (OR 0.3, 95% CI 0.11-0.81), and physical activity (2.8 95% CI 1.0-7.8), but mindfulness scores did not change from baseline to follow-up.

Conclusions: The Headspace mobile mindfulness app was easy and cost-effective to implement and acceptable to those who participated, but few women elected to try it. The unique characteristics of this southern Louisiana population suggest that more intense promotion of the benefits of mindfulness training is needed, perhaps in conjunction with some therapist or researcher support. Several short-term benefits of the app were identified, particularly for depression and sleep.

(JMIR Mhealth Uhealth 2020;8(6):e15943) doi:10.2196/15943

KEYWORDS

mindfulness; mobile phone; depressive symptoms; women; Louisiana
Introduction

Background

Mindfulness refers to a state of consciousness that focuses on an individual’s attention and awareness in the present moment [1]. Mindfulness-based stress reduction (MBSR) is a standardized meditation program created from efforts to integrate Buddhist mindfulness meditation with contemporary clinical and psychological practice [2]. MBSR and other forms of mindfulness-based therapy have been observed in systematic reviews and meta-analyses to convey a variety of beneficial mental health outcomes, such as lowered anxiety, stress, and depression [3-5]. Other studies have also explored its effect on sleep quality [3,6,7], physical activity [8,9], smoking [9], and eating behaviors [9,10], although with mixed or inconclusive results.

Traditional MBSR programs can be resource intensive for providers and time- and cost-intensive for participants. Interventions typically consist of 8 weekly meetings led by a trained facilitator, with daily homework exercises and a weeklong retreat [2]. Individuals likely to benefit from this type of intervention may not have the time or resources to engage in such programs nor have easy access to experienced leaders. Thus, alternative low-intensity self-help methods are needed to expand the reach of traditional mindfulness-based approaches. Given that 77% of Americans now own a smartphone [11], the use of mobile technologies for this purpose may be particularly convenient and cost-effective for populations that are busy, less affluent, or geographically distant from skilled providers. Low-intensity [12] and web-based interventions [13] have begun to show promise in improving mindfulness, depression, and stress outcomes.

Southern Louisiana is a vulnerable, disaster-prone region with a highly stressed population. Recent disasters, such as the 2005 hurricanes and the 2010 Deepwater Horizon oil spill, have resulted in significant cumulative mental health impacts related to depression, anxiety, psychological distress, and posttraumatic stress disorder [14-17]. Mental health services were decimated following the 2005 hurricanes [16,18], resulting in residents receiving less mental health treatment than they required. Women, in particular, represent an influential yet vulnerable and understudied population. They are often central to decision-making processes within households, particularly with respect to decisions regarding health, support, diet, and caregiving. Therefore, there is a need for low-cost, easy-to-implement mental health and healthy lifestyle interventions that can be disseminated over large population areas. As MBSR has shown promising results, particularly in the area of stress reduction [5], a low-intensity mindfulness-based intervention may be a useful tool in the disaster recovery toolbox.

Objectives

This study was designed to assess the use of a mobile mindfulness program under real-world conditions within a community sample of women residing in southern Louisiana. The objectives of this study were to (1) assess the feasibility and acceptability of a mobile mindfulness meditation program, (2) describe the predictors of program usage among study participants, and (3) assess the effect of the program on secondary health outcomes.

Methods

Participants

Participants of the Women and Their Children’s Health (WaTCH) study, designed to investigate the physical and mental health effects of the Deepwater Horizon oil spill in Louisiana, were invited to enroll in the study. WaTCH participants (n=2852) were followed over 2 waves of data collection (2012-2016). Eligible women were aged 18 to 80 years at the time of initial data collection from 2012 to 2014 and resided in 1 of 7 parishes in southern Louisiana on April 20, 2010 [19]. A total of 1376 adult WaTCH participants who provided valid email addresses were invited to participate in a prospective pilot study on MBSR and stress among women during the summer of 2017. Of 1376 participants, 526 consented to and completed the baseline survey, of whom 318 consented to participate in the MBSR mobile mindfulness app component of the study, and 236 completed the follow-up survey and comprised the final sample. Figure 1 shows the study flow chart. The study was approved by the Louisiana State University Health Sciences Center (LSUHSC) New Orleans institutional review board.
Procedure

Recruitment
Adult WaTCH participants were recruited through email invitation, consented through the web, and administered a web-based baseline survey via Research Electronic Data Capture hosted at the Epidemiology Data Center at the LSUHSC School of Public Health [20]. On completion of the baseline survey, subjects were asked to use a mobile mindfulness app (Headspace, described under The Headspace Program) and asked to indicate their consent by signing a data use privacy acknowledgment form. They were then given a redemption code for free access to the app. During the study period, subjects had no interaction with the study team. Headspace has its own built-in reminders that can be programmed according to user specifications. After 45 days, the subjects were prompted to complete the follow-up survey. Subjects completed the follow-up survey an average of 80.7 (SD 49.2) days after the baseline survey.

Email Reminders
At each subsequent step of the study (ie, baseline survey consent, baseline survey, Headspace acknowledgment, and follow-up survey), those who had not completed the task were sent periodic automated reminders every 4 days until task completion, up to 3 times. For the follow-up survey, automated reminders were sent up to 5 times. A one-time email reminder was also sent automatically 7 days after the Headspace acknowledgment was signed to remind the subjects to complete the program. The study staff had no information on whether subjects had downloaded or used the app until after study completion.

Incentives
Subjects were given US $10 on completion of the baseline survey and US $10 on completion of the follow-up survey. Those who completed the Headspace acknowledgment form were also given free access to Headspace for 1 year (value approximately US $96).

The Headspace Program
The mobile MBSR training program consisted of using a smartphone or web-based app called Headspace. Headspace was selected because it had the highest average rating in a review of 23 apps that provided mindfulness training and education [21]. Subjects were asked to download the app and use it for at least 30 days, 10 min at a time. After 45 days, they were reminded to complete a web-based follow-up survey. The app contained a Foundations series consisting of 3 groups of 10 sessions at the Basics level as well as a variety of other themed packs. Most sessions were designed to be used for 10 min per day. Participants had the ability to explore the app and complete any of the other sessions they desired for a full year. Although the study staff had no formal contact with participants during the study period, they were available to provide technical
support on request (via an email or a toll-free number). On completion of the data collection period, Headspace developers provided the researchers with participant data on number of sessions, date/time of meditation sessions attempted, and platform uses.

Measures

The baseline and follow-up surveys consisted of questions about participant demographics, physical and mental health, and social and environmental characteristics. The follow-up survey also contained questions related to acceptability of the Headspace app program.

**Primary Outcomes**

The primary outcomes of the study were feasibility, acceptability, and characteristics of app utilization. Feasibility was defined as (1) enrollment (eligible subjects who consented to the study and completed the Headspace acknowledgment form), (2) program participation (enrolled subjects who logged into Headspace at least once during the study period), and (3) retention (enrolled subjects who completed the follow-up survey). Acceptability was measured through a series of closed-ended questions about how well participants liked the app. Example questions included “How would you rate the Headspace app on a scale of 1 to 5?” “Would you recommend Headspace to others?” and “How much of the Headspace program did you complete?” All subjects were also asked, “What do you think were the biggest barriers to completing the Headspace program?” Example responses that subjects were allowed to choose from included “not enough time,” “not interested in mindfulness meditation,” “didn’t see how mindfulness meditation would benefit me,” “no privacy or quiet space to do the meditation,” “did not like the guy’s voice,” “didn’t have access to a smartphone or computer every day,” and “technical problems.” Subjects were allowed to select as many responses as applied, including an option to specify something different. Data on characteristics of Headspace app usage included the total number of log-ins to Headspace, average log-ins per program completer, platform used (iOS, Android, or web-based), day of week of use (weekday vs weekend), and time of day of use (in 4-hour blocks).

**Secondary Outcomes**

The secondary outcomes of the study included trait mindfulness, depressive symptoms, perceived stress, sleep quality, physical activity, body mass index (BMI), and healthy eating.

**Mindfulness**

Mindfulness was measured using the 15-item Mindful Attention Awareness Scale (MAAS), trait version, and ranged from 1=almost always to 6=almost never [1]. MAAS scores were averaged and dichotomized at the median.

**Depressive Symptoms**

Depressive symptoms were measured for the past week with the Center for Epidemiologic Studies Depression Scale-10 (CESD-10) [22]. Respondents rated the frequency of symptoms that occurred during the past week on a 4-point scale, ranging from 0=none of the time to 3=most of the time. Item scores were summed after reverse coding the positive mood items (range 0-30). They were then dichotomized such that total scores of 10 or greater were indicative of depressive symptoms [22].

**Perceived Stress**

Perceived stress was measured using the Perceived Stress Scale, 4-item version (PSS-4) [23], with responses ranging from 0=never to 4=very often. Items were summed after reverse scoring the positively worded items (range 0-16), with higher scores indicating greater levels of stress. PSS-4 scores for this study were dichotomized at the median.

**Sleep Quality**

Sleep quality was measured using the Pittsburgh Sleep Quality Index (PSQI), a 19-item self-rated questionnaire that assesses sleep quality and disturbances over a 1-month time interval [24]. A total of 7 component scores were created: subjective sleep quality, sleep latency (length of time it takes to fall asleep), sleep duration, habitual sleep efficiency (the percentage of time in bed that one is asleep), sleep disturbances, use of sleeping medication, and daytime dysfunction. The sum of the component scores yielded one global score (range 1-19), where higher scores indicate more sleep problems. A global score ≤5 is indicative of good sleep quality. Responses to the individual components of the index were grouped into 4 categories, ranging from 0 (better) to 3 (worse), and then dichotomized into 2 groups.

**Physical Activity**

Physical activity was measured using the Total Physical Activity Screener from the Stanford Brief Activity Survey [25]. In total, 2 questions asked about on-the-job activity and leisure-time activity. Responses were categorized into 5 levels of physical activity, from inactive to very hard intensity, and further dichotomized into inactive/light-intensity and moderate/hard/very hard–intensity activity.

**Body Mass Index**

BMI was calculated from self-reported height and weight, using the formula of weight in kilograms divided by height in meters squared, and then grouped into 2 groups: normal/underweight versus overweight/obese.

**Healthy Eating**

Healthy eating was assessed using items from the Dietary Screener of the 2009 California Health Interview Survey [26], which gathers information about the intake of fruits and vegetables and teaspoons of added sugar. In total, 2 summary measures were calculated: daily cup equivalents of fruits and vegetables and daily teaspoons of added sugar. Each measure was dichotomized at the median.

**Other Covariates**

Age at the time of the interview, race/ethnicity, household income, marital status, employment status, and number of minor children living in the household were also measured.

**Analysis**

Descriptive statistics were calculated for all measures. Feasibility was assessed by calculating the enrollment percentage, program participation, and retention. Characteristics of those who consented to the program (N=318) were compared
with those who completed both surveys (N=236). Baseline characteristics and per-protocol results were assessed. Comparisons of secondary outcomes between program participants and nonparticipants were performed using Pearson chi-square or Fisher exact tests for categorical variables. Multivariable logistic regression modeling was used to assess the associations between the secondary outcomes at follow-up and participation in the Headspace program. When needed, Firth penalized logistic regression models were used to overcome separation issues [27]. As the corresponding outcome at baseline and the total number of days each subject used the Headspace app were identified as potential confounders for the majority of secondary outcomes (ie, the crude and adjusted measures of association differed appreciably [28]), all regression models were adjusted for both variables. All statistical tests were carried out using SAS 9.3 (SAS Institute Inc) at a type 1 error level of 0.05.

Results

Feasibility

Of 526 women who were eligible to participate in the program because they had completed the baseline survey, 318 consented and completed the Headspace acknowledgment form, resulting in an enrollment of 60.5% of the eligible sample. Of the 318 women who enrolled, 43 (13.5%) actually participated in the program. Of those who enrolled, 236 women completed the follow-up survey, yielding a retention proportion of 74.2%.

Baseline Characteristics of the Sample

Table 1 shows the demographic characteristics of the women in the sample. Women who consented to the study were similar to women who consented and completed follow-up measures. Most women who completed follow-up measures were white (140/236, 59.3%), had less than a college education (144/236, 61.0%), had a household income of less than US $50,000 per year (119/236, 50.4%), and were currently working full time or part time (156/236, 66%). Women who participated in the program were more likely to have a college education (27/43, 63% vs 65/193, 33.6%; \( P < .001 \)) and be currently working outside the home (34/43, 79% vs 122/193, 63.2%; \( P = .047 \)) than women who did not participate.

Table 1. Baseline demographic characteristics of the sample by program participation, Louisiana, 2017 to 2018.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total samplea (N=236)</th>
<th>Program participantsa (N=43)</th>
<th>Nonparticipantsa (N=193)</th>
<th>( P ) value</th>
<th>Consent onlya (N=318)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race/ethnicityb, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.19</td>
<td>N/Ac</td>
</tr>
<tr>
<td>Non-Hispanic white</td>
<td>140 (59.3)</td>
<td>29 (67.4)</td>
<td>111 (57.5)</td>
<td>N/A</td>
<td>165 (51.8)</td>
</tr>
<tr>
<td>Non-Hispanic black or other/multi/Hispanic</td>
<td>93 (39.4)</td>
<td>13 (30.2)</td>
<td>80 (41.4)</td>
<td>N/A</td>
<td>137 (43.0)</td>
</tr>
<tr>
<td>Educationc, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td>N/A</td>
</tr>
<tr>
<td>High school graduate or less</td>
<td>144 (61.0)</td>
<td>16 (37.2)</td>
<td>128 (66.3)</td>
<td>N/A</td>
<td>195 (61.3)</td>
</tr>
<tr>
<td>College or more</td>
<td>92 (38.9)</td>
<td>27 (62.7)</td>
<td>65 (33.6)</td>
<td>N/A</td>
<td>122 (38.3)</td>
</tr>
<tr>
<td>Current household incomeb, US $, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>N/A</td>
</tr>
<tr>
<td>( \leq 50,000 ) per year</td>
<td>119 (50.4)</td>
<td>17 (39.5)</td>
<td>102 (52.8)</td>
<td>N/A</td>
<td>165 (51.8)</td>
</tr>
<tr>
<td>( &gt; 50,000 ) per year</td>
<td>100 (42.3)</td>
<td>23 (53.4)</td>
<td>77 (39.8)</td>
<td>N/A</td>
<td>134 (42.1)</td>
</tr>
<tr>
<td>Marital status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.32</td>
<td>N/A</td>
</tr>
<tr>
<td>Married or living with a partner</td>
<td>149 (63.1)</td>
<td>30 (69.7)</td>
<td>119 (61.6)</td>
<td>N/A</td>
<td>203 (63.8)</td>
</tr>
<tr>
<td>Widowed, divorced, separated, or never married</td>
<td>87 (36.8)</td>
<td>13 (30.2)</td>
<td>74 (38.3)</td>
<td>N/A</td>
<td>115 (36.1)</td>
</tr>
<tr>
<td>Employment statusb, n (%)</td>
<td></td>
<td></td>
<td></td>
<td>.047</td>
<td>N/A</td>
</tr>
<tr>
<td>Currently working full time or part time</td>
<td>156 (66.1)</td>
<td>34 (79.0)</td>
<td>122 (63.2)</td>
<td>N/A</td>
<td>211 (66.3)</td>
</tr>
<tr>
<td>Not currently working full time or part time</td>
<td>80 (33.8)</td>
<td>9 (20.9)</td>
<td>71 (36.7)</td>
<td>N/A</td>
<td>107 (33.6)</td>
</tr>
<tr>
<td>Ageb (years), mean (SD)</td>
<td>46.1 (10.0)</td>
<td>46.6 (9.8)</td>
<td>46.0 (10.1)</td>
<td>.82</td>
<td>46.8 (10.3)</td>
</tr>
<tr>
<td>Number of children aged &lt;18 years living in a house-</td>
<td>1.1 (1.2)</td>
<td>0.8 (0.9)</td>
<td>1.2 (1.3)</td>
<td>.18</td>
<td>1.1 (1.2)</td>
</tr>
<tr>
<td>hold, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aTotal sample (N=236) includes those who completed both the baseline and follow-up surveys. Program participants (N=43) include program completers, those who logged into the Headspace app at least once and completed both surveys. Nonparticipants (N=193) include program noncompleters. Consent only (N=318) includes those who completed the baseline survey and consented to the program but did not complete the follow-up survey.

bRace/ethnicity missing (n=3); income missing (n=17); and age missing (n=1).

cN/A: not applicable.
Acceptability of Headspace App

The acceptability of the program was assessed among the 43 women who used the app and subsequently completed the follow-up survey (Table 2). Most log-ins (1191/1530, 77.8%) took place on an iOS device and on a weekday (1147/1530, 75.0%). Most sessions (375/1530, 24.5%) were conducted in the afternoons between noon and 4 PM, and another 20.3% (310/1530) of sessions were conducted in the early evenings between 4 PM and 8 PM. Women logged into the app, on average, 36 times (SD 80) for an average of 24 days (SD 36). This includes one enthusiastic participant who logged in a total of 503 times and another who logged in over 156 days.

Participants who used the Headspace app were also asked about their experiences with it. Almost three-fourth of participants reported that they were pleased with it or loved it, and more than 85% said they would recommend the app to others. More than two-third of participants reported that they had completed at least some or all of the program. More than two-third of participants said they liked the relaxation aspect of Headspace the best, followed by the voice of the meditation leader, and the duration of the session.

Participants (N=43) and nonparticipants (N=193) were also asked to report what they did not like about the Headspace app (Table 3). The biggest barriers cited among those who managed to participate were lack of time (16/43, 37%), lack of privacy (8/43, 19%), and lack of a quiet space to meditate (8/43, 19%). Among those who did not participate, the biggest reported barriers were time (94/193, 48.7%), lack of a quiet space (26/193, 13.5%), and lack of privacy (21/193, 10.9%).
Table 2. Characteristics of Headspace usage among participants.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Access characteristics per log-in (N=1530), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Platform used</strong></td>
<td></td>
</tr>
<tr>
<td>iOS</td>
<td>1191 (77.8)</td>
</tr>
<tr>
<td>Android</td>
<td>116 (7.6)</td>
</tr>
<tr>
<td>Desktop</td>
<td>223 (14.6)</td>
</tr>
<tr>
<td><strong>Day of week log-in session occurred</strong></td>
<td></td>
</tr>
<tr>
<td>Weekday</td>
<td>1147 (75.0)</td>
</tr>
<tr>
<td>Weekend</td>
<td>383 (25.0)</td>
</tr>
<tr>
<td><strong>Time of day log-in session began</strong></td>
<td></td>
</tr>
<tr>
<td>Midnight to 4 AM</td>
<td>301 (19.7)</td>
</tr>
<tr>
<td>4 AM to 8 AM</td>
<td>243 (15.9)</td>
</tr>
<tr>
<td>8 AM to noon</td>
<td>158 (10.3)</td>
</tr>
<tr>
<td>Noon to 4 PM</td>
<td>375 (24.5)</td>
</tr>
<tr>
<td>4 PM to 8 PM</td>
<td>310 (20.3)</td>
</tr>
<tr>
<td>8 PM to midnight</td>
<td>143 (9.3)</td>
</tr>
<tr>
<td><strong>Access per participant (N=43), mean (SD; range)</strong></td>
<td></td>
</tr>
<tr>
<td>Log-ins to Headspace</td>
<td>35.6 (80.3; 1-503)</td>
</tr>
<tr>
<td>Number of days used Headspace</td>
<td>24.0 (36.1; 1-156)</td>
</tr>
<tr>
<td><strong>Acceptability per participant (N=43), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Overall rating of Headspace app</td>
<td></td>
</tr>
<tr>
<td>Hated it</td>
<td>1 (2.3)</td>
</tr>
<tr>
<td>Not crazy about it</td>
<td>6 (14.0)</td>
</tr>
<tr>
<td>Feel neutral about it</td>
<td>4 (9.3)</td>
</tr>
<tr>
<td>Pleased with it</td>
<td>19 (44.2)</td>
</tr>
<tr>
<td>Loved it</td>
<td>13 (30.2)</td>
</tr>
<tr>
<td>Would recommend Headspace app to others</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>37 (86.1)</td>
</tr>
<tr>
<td>No</td>
<td>6 (14.0)</td>
</tr>
<tr>
<td><strong>How much of Headspace completed</strong></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>A little</td>
<td>12 (27.9)</td>
</tr>
<tr>
<td>Some</td>
<td>15 (34.9)</td>
</tr>
<tr>
<td>All</td>
<td>16 (37.2)</td>
</tr>
<tr>
<td><strong>What did you like best about the Headspace program?</strong></td>
<td></td>
</tr>
<tr>
<td>Relaxation</td>
<td>12 (36.4)</td>
</tr>
<tr>
<td>Voice</td>
<td>5 (15.2)</td>
</tr>
<tr>
<td>Good length of time</td>
<td>4 (12.1)</td>
</tr>
<tr>
<td>Good concept</td>
<td>3 (9.1)</td>
</tr>
<tr>
<td>Forced me to take personal time</td>
<td>3 (9.1)</td>
</tr>
<tr>
<td>Easy access</td>
<td>2 (6.1)</td>
</tr>
<tr>
<td>Completed</td>
<td>1 (3.0)</td>
</tr>
<tr>
<td>Slept better</td>
<td>1 (3.0)</td>
</tr>
</tbody>
</table>
Effect of the Headspace Program on Secondary Outcomes

A description of the secondary outcomes at baseline is shown in Multimedia Appendix 1. Data are presented for the total sample of those who completed both surveys (N=236), for the program participants (N=43) and nonparticipants (N=193), and for those who consented but did not complete the program or the follow-up survey. At baseline, almost half of the study sample (111/227, 48.9%) reported high levels of depressive symptoms, 59.8% (141/236) reported high levels of perceived stress, 64.7% (141/218) reported high levels of sleep problems, 81.7% (183/224) reported BMI in the overweight or obese category, and 50.3% (94/187) reported high levels of sugar intake. In terms of healthy behaviors, 50.4% (115/227) of the study sample reported high levels of mindfulness, 36.8% (86/234) reported high levels of physical activity, and 50.0% (114/228) reported high levels of daily fruit and vegetable intake. Program participants and nonparticipants tended to have similar scores on all outcomes at baseline, as did the group that only consented.

The results from logistic regression models predicting the effect of Headspace program participation on secondary outcomes at follow-up are shown in Table 4. All models were adjusted for condition at baseline and the total number of days the app was used. Depressive symptoms, sleep quality, sleep duration, sleep latency, physical activity, and fruit and vegetable intake all improved after participation in the Headspace program. Those who participated in the Headspace program were 0.3 (95% CI 0.11-0.81) times likely to be depressed at follow-up, 0.1 (95% CI 0.02-0.96) times likely to have poor sleep quality, 0.3 (95% CI 0.07-0.86) times likely to have poor sleep duration, 0.3 (95% CI 0.12-0.99) times likely to have poor sleep latency, 2.8 (95% CI 1.0-7.8) times likely to participate in moderate to very hard physical activity, and 0.94 (95% CI 0.99-5.78) times likely to have increased fruit and vegetable intake. No changes at follow-up were observed for mindfulness or other health indicators.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Participants (N=43), n (%)</th>
<th>Nonparticipants (N=193), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not enough time</td>
<td>16 (37.2)</td>
<td>94 (48.7)</td>
</tr>
<tr>
<td>No privacy to do the meditation</td>
<td>8 (18.6)</td>
<td>21 (10.9)</td>
</tr>
<tr>
<td>No quiet space to do the meditation</td>
<td>8 (18.6)</td>
<td>26 (13.5)</td>
</tr>
<tr>
<td>Did not like the guy's voice on the Headspace app</td>
<td>7 (16.3)</td>
<td>7 (3.6)</td>
</tr>
<tr>
<td>Not interested in mindfulness meditation</td>
<td>3 (7.0)</td>
<td>19 (9.8)</td>
</tr>
<tr>
<td>Didn’t see how mindfulness meditation would benefit me</td>
<td>3 (7.0)</td>
<td>16 (8.3)</td>
</tr>
<tr>
<td>Technical problems with installation or use</td>
<td>0 (0.0)</td>
<td>12 (6.2)</td>
</tr>
<tr>
<td>Didn’t have access to a smartphone or computer every day</td>
<td>1 (2.3)</td>
<td>17 (8.8)</td>
</tr>
</tbody>
</table>

Table 3. Barriers to Headspace use.
Table 4. Individual logistic regression models predicting the effect of the program on outcome at follow-up, adjusted for number of days app was used, and outcome at baseline.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants, N</th>
<th>Odds ratio (95% Wald confidence limits)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater mindfulness (MAAS(^a) ≥4.13)</td>
<td>219</td>
<td>1.69 (0.53-5.38)</td>
<td>.37</td>
</tr>
<tr>
<td>More depressive symptoms (CESD-10(^b) ≥10)</td>
<td>224</td>
<td>0.29 (0.11-0.81)</td>
<td>.02</td>
</tr>
<tr>
<td>Greater perceived stress (PSS(^c) ≥6)</td>
<td>235</td>
<td>0.76 (0.31-1.85)</td>
<td>.55</td>
</tr>
<tr>
<td><strong>PSQI(^d)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor habitual sleep efficiency</td>
<td>196</td>
<td>1.56 (0.52-4.64)</td>
<td>.43</td>
</tr>
<tr>
<td>Poor sleep quality</td>
<td>196</td>
<td>0.14 (0.02-0.96)</td>
<td>.045</td>
</tr>
<tr>
<td>Need for medications to sleep</td>
<td>196</td>
<td>0.47 (0.13-1.70)</td>
<td>.25</td>
</tr>
<tr>
<td>Poor sleep duration</td>
<td>196</td>
<td>0.25 (0.07-0.86)</td>
<td>.03</td>
</tr>
<tr>
<td>Sleep disturbance</td>
<td>196</td>
<td>1.04 (0.41-2.68)</td>
<td>.93</td>
</tr>
<tr>
<td>Poor sleep latency</td>
<td>196</td>
<td>0.34 (0.12-0.99)</td>
<td>.048</td>
</tr>
<tr>
<td>Day dysfunction due to sleepiness</td>
<td>196</td>
<td>0.44 (0.14-1.41)</td>
<td>.17</td>
</tr>
<tr>
<td>Total PSQI score &gt;5</td>
<td>196</td>
<td>2.35 (0.63-8.77)</td>
<td>.20</td>
</tr>
<tr>
<td><strong>Physical activity (moderate/hard/very hard intensity)</strong></td>
<td>221</td>
<td>2.79 (1.00-7.78)</td>
<td>.05</td>
</tr>
<tr>
<td><strong>BMI(^e)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(overweight/obese)</td>
<td>216</td>
<td>0.52 (0.06-4.67)</td>
<td>.56</td>
</tr>
<tr>
<td><strong>Healthy eating measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit and vegetable intake ≥0.91 daily cup equivalents</td>
<td>201</td>
<td>0.94 (0.99-5.78)</td>
<td>.05</td>
</tr>
<tr>
<td>Sugar intake ≥6.25 teaspoons</td>
<td>160</td>
<td>1.00 (0.35-2.86)</td>
<td>.99</td>
</tr>
</tbody>
</table>

\(^a\)MAAS: Mindful Attention Awareness Scale.
\(^b\)CESD-10: Center for Epidemiologic Studies Depression Scale-10.
\(^c\)PSS: Perceived Stress Scale.
\(^d\)PSQI: Pittsburgh Sleep Quality Index.
\(^e\)BMI: body mass index.
\(^f\)Firth penalized logistic regression model used to overcome separation issues.

Discussion

The objectives of this study were to (1) assess the feasibility and acceptability of a mobile mindfulness meditation program among women residing in southern Louisiana, (2) describe the predictors of program usage among study participants, and (3) assess the effect of the program on secondary health outcomes.

Feasibility, Acceptability, and Predictors of App Usage

Retention

In this study, 74% (32/43) of the participants completed the follow-up survey. A systematic review and meta-analysis of mindfulness self-help interventions, including delivery through the web, reported that an average of 73% of randomized participants completed posttreatment measures, which is comparable with attrition in studies of other self-help and minimal contact therapies [12]. This study demonstrated similar findings, suggesting that it is, although not a randomized trial, within the general norms for study completion.

Program Participation

However, only 13.5% (43/318) of participants in this study actually participated in the program, which was broadly defined as logging in to the Headspace app at least once. This is substantially lower than the participation rates demonstrated in similar studies. A review of self-help mindfulness intervention randomized controlled trials (RCTs) found that 48% of participants met the study-defined intervention engagement criteria [12]. Similarly, over a dozen studies to date have been published evaluating the feasibility or effectiveness of Headspace, and many of these also have stronger participation rates than this study. This calls into question whether Headspace is really a feasible intervention for this particular population of women. The majority of Headspace studies employed samples of university students or residents [29-35], samples of employees [36,37], or clinical samples [38,39]. Only 3 studies of Headspace used community samples, and these participants were predominantly self-selecting, white, well-educated [40,41], and living in Australia [40] or the United Kingdom [42]. This contrasts with the sample of this study, which was recruited from a cohort of women representative of the area, of whom 35% were African Americans, only 39% had a college education or higher, and which took place in southern coastal Louisiana, an area that has been subject to quite a few natural and man-made disasters in recent years. Although our sample may have been more highly stressed than national norms [43], they were not recruited into the study to address any particular...
clinical condition nor were they offered any suggestions that the program would help them with their own issues, as might happen with a sample of participants recruited from a clinic. Thus, their perceived need to engage with the app may have been less. In addition, evidence suggests that clinical or therapist support is beneficial for promoting adherence in behavioral web-based interventions [12,44]. Prior studies of Headspace incorporated some in-person contact between participants and study personnel, either in the form of an initial briefing session [29,30,37] or by completing some sessions in a group setting [32,34]. However, most of these studies were interested in taking advantage of the completely self-directed nature of Headspace and learning how effective the app could be when administered entirely on the web with little engagement from study personnel. It is possible that this particular population, consisting of a less educated and significantly minority population, will require more intense personal encouragement to persuade them to use the app. Future interventions could consider including some type of nonautomated or in-person support. Another possibility for the low level of engagement may be because of study fatigue. Since 2012 and before this study, WaTCH participants had been asked to take part in two telephone surveys and a home visit, not to mention other research studies that were taking place in the area at the same time.

**Barriers**

The main barrier to using the Headspace app was lack of time, cited by both participants and nonparticipants alike, a finding that is echoed in the literature. A qualitative study of Headspace users found that the main concern of users was fitting the app into their busy lives [45], whereas a study of another wellness mobile app reported that being busy made it difficult to find suitable, peaceful moments to engage in the intervention exercises [46]. Mobile-based interventions, by virtue of their ability to be used anywhere at any time, are designed to address this concern, but competing demands on participants’ time make this a continuing challenge.

**Predictors of App Usage and Acceptability**

Differences between participants and nonparticipants may shed light on who is most likely to adopt the program and thus who the program is most likely to benefit. Those who chose to engage with the program tended to be more highly educated, have higher incomes, and be employed compared with those who did not, characteristics similar to those found in a national study of mindfulness practices [47]. It may be that women with these characteristics are more familiar with meditation practices in general and therefore more likely to use the Headspace app. Alternatively, it may be that the more educated and higher-income women are more likely to use mobile devices, as ownership of smartphones increases with education and income [11]. As the use of mobile devices is likely to keep increasing, additional promotion of the benefits of mindfulness practice may be warranted. Women who chose to use the app also tended to be less stressed. This suggests that those who perhaps need the program the most will require more persuasion to get them to use it.

Nevertheless, those who engaged with the app did so at high levels, self-reporting the completion of some if not all the sessions as well as high levels of satisfaction with the app. These findings highlight the importance of understanding the characteristics of those who choose to use the app [48], suggesting that it is an acceptable program for reaching large numbers of individuals.

**Secondary Outcomes**

This study also explored the impact of mobile mindfulness-based program participation on a number of secondary health and behavioral outcomes. Participation had a beneficial impact on depressive symptoms, sleep indicators, and physical activity, but mindfulness scores did not change from baseline to follow-up.

**Depression**

Almost half of the women in our community sample reported symptoms of depression. A study among older adults, using the same scale and cutoff values, demonstrated the prevalence of depressive symptoms between 12.3% and 16.3% [22], providing more evidence that our own population may be highly stressed. Subjects who used the Headspace app were less likely to experience depressive symptoms at follow-up than women who did not use it. Most of the comprehensive reviews of depression and MBSR conducted in clinical samples using noninternet-based interventions [3,4,49,50] have demonstrated improvements in depression following MBSR, and studies in both clinical and community samples using internet- or mobile-based delivery have also noted decreases in depressive symptoms [37,51-53]. That we were able to detect similar improvements in our community sample, although uptake of the app was low, suggests that this may be a viable program to improve depressive symptoms in this population.

**Sleep**

A large proportion of our study participants reported sleep problems, particularly in the areas of sleep latency and sleep disturbance. Population-based studies of insomnia suggest that approximately 30% of the general population has complaints of sleep disruption, with female gender and older age being the predominant demographic risk factors [54]. Depression is one of the most common comorbid psychiatric disorders in insomniacs [55], whereas other consequences of sleep difficulties include decreased quality of life, increased accidents, and reduced work productivity [55]. Given the high prevalence of depression and sleep problems in our sample, the search for effective interventions could be critical. In our study, 3 components of the PSQI showed improvement after participants used the Headspace app: self-reported sleep quality, sleep duration, and sleep latency. The evidence for a beneficial effect of mindfulness programs on sleep in the literature is inconclusive. One review of studies found that MBSR significantly improved measures of sleep quality or duration [6], but 2 later reviews either found insufficient [3] or mixed evidence [7]. However, none of these reviews looked specifically at internet-based programs. Our results suggest that using Headspace may be beneficial for sleep, but more research in larger populations is warranted.
Physical Activity

Physical Activity Guidelines for Americans recommends at least 150 min per week of moderate-intensity physical activity [56], yet fewer than half of our sample reported being in at least the moderate-intensity category over the past year. Those who used the app were borderline more likely to see improvements in their physical activity levels after the program than those who did not. This is consistent with a systematic review of RCTs finding that mindfulness training had a positive effect on physical activity levels [57] and a review of cross-sectional studies indicating positive relationships between dispositional mindfulness and physical activity [58]. Moreover, results from the 2012 National Health Interview Survey showed that US adults who practiced mindfulness meditation in the past year were less likely to be inactive and more likely to meet physical activity recommendations [59]. The findings from this study indicate that using the Headspace app may be beneficial for physical activity, but more research in a larger population is warranted.

Mindfulness

MBSR is thought to improve certain health outcomes through its improvement of individuals’ levels of dispositional mindfulness [60-62]. A positive relationship between the program and levels of mindfulness would be a first step in testing whether this mechanism improves health. In this study, however, despite experiencing improvement in some secondary outcomes, participants did not significantly improve their mindfulness scores. Evidence in the scientific literature that self-reported mindfulness is a primary mechanism of change is mixed [63,64], and some have questioned the validity of these measures, including the MAAS used in this study [65]. Among the arguments posed are the lack of a clear gold standard with which to define a mindful person, the lack of consensus about what mindfulness is, and a debate about whether individuals can accurately self-report their own levels of mindfulness [65]. The MAAS instrument used to assess mindfulness in this study is a well-regarded and validated instrument [1], yet our results call into question whether self-reported mindfulness is indeed the actual mechanism through which the program improves health outcomes. Other potential reasons for the lack of significant associations between the program and mindfulness may be related to differences in duration and dosage of the intervention, characteristics of our sample population, and differences in levels of motivation to participate.

Conclusions

This study was designed to assess the use of a mobile mindfulness intervention in real-world conditions in a community sample of women residing in southern Louisiana. The all-electronic program was easy and cost-effective to implement and acceptable to those who participated, but few women elected to try it. The unique characteristics of this population (more prone to environmental disasters, more minorities, less educated, and lower income) suggest that more intense promotion of the benefits of mindfulness training is needed, perhaps in conjunction with some therapist or researcher support. Several short-term benefits of the program were identified, particularly for depression and sleep.

Limitations

Some limitations of the study were identified: (1) No control group was used, making it difficult to determine whether the observed effects were because of the use of the app or other characteristics of the sample that led them to choose to engage with the app. It is possible that participants who already believed in the benefits of meditation were more likely to choose to use the app and to report that they benefited from it than those who did not. (2) The length of time spent in mindfulness sessions was not clearly defined in this study. Participants were asked to spend at least 10 min per day using the app, but the researchers did not have access to specific data reporting the exact length of time each participant actually spent with the app. More information regarding dosage would help inform how much mindfulness exposure is needed to achieve benefits. (3) Participants were asked to use the app only for a period of 30 days, although they had access to it for much longer. Although we adjusted for the number of days participants used the app, meditation is a skill that many spend a lifetime learning and that traditional MBSR programs teach for longer periods. The length of exposure in this study may not have been sufficient to detect changes in secondary outcomes. (4) The sample was exclusively female, limiting the generalizability to men. (5) The sample differed from the original WaTCH cohort in that they were slightly younger and were more likely to be college educated, have a higher income, and be working full time than the original cohort. (6) The incentive structure of the study resulted in subjects being paid for completing both surveys, but not being paid for participating in the actual mindfulness program itself. This may have led to a potential bias in the composition of the participant and nonparticipant groups, as participants tended to be better educated and employed than nonparticipants.

Strengths

The strengths of this study include (1) the setting of the program in southern Louisiana, within a rural population residing in a low-density area where access to group-based in-person mindfulness approaches is sparse and likely not sustainable; (2) the population was a sample of women residing in the community, without specific clinical manifestations, and thus addresses the need to translate prior research into effective and sustainable community mindfulness intervention programs [63]; (3) the sample contained individuals from a variety of racial, educational, and socioeconomic backgrounds that are traditionally underrepresented among mindfulness practitioners. As the study was not an RCT, it was able to highlight some important differences between individuals who chose to engage with the app and those who did not, namely, in terms of income, education, and stress levels; (4) the study used mobile technology to implement a self-help program approach to mindfulness, addressing issues such as lack of access to facilities and instructors, resources, and time to meditate.

This study demonstrated the feasibility and acceptability of implementing a mobile mindfulness meditation program in a population of women in southern Louisiana. Such apps represent a convenient and cost-effective technology that can easily be scaled up to address barriers to the implementation of traditional
MBSR programs but may require supplemental support to promote their use.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Baseline measurement of secondary outcomes by program participation, Louisiana, 2017 to 2018. [DOCX File, 20 KB - mhealth_v8i6e15943_app1.docx]

References


Abbreviations

CESD-10: Center for Epidemiologic Studies Depression Scale-10
LSUHSC: Louisiana State University Health Sciences Center
MAAS: Mindful Attention Awareness Scale
MBSR: mindfulness-based stress reduction
OR: odds ratio
PSQI: Pittsburgh Sleep Quality Index
PSS-4: Perceived Stress Scale, 4-item version
RCT: randomized controlled trial
WaTCH: Women and Their Children’s Health
Effectiveness of a 3-Month Mobile Phone–Based Behavior Change Program on Active Transportation and Physical Activity in Adults: Randomized Controlled Trial

Anna Ek, PhD; Christina Alexandrou, MSc; Emmie Söderström, MSc; Patrick Bergman, PhD; Christine Delisle Nyström, PhD; Artur Direito, PhD; Ulf Eriksson, PhD; Pontus Henriksson, PhD; Ralph Maddison, PhD; Ylva Trolle Lagerros, PhD; Marcus Bendtsen, PhD; Marie Löf, PhD

Department of Clinical Science Intervention and Technology, Karolinska Institutet, Huddinge, Sweden
Department of Health, Medicine and Caring Sciences, Linköping, Sweden
Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden
Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
Strömmstad Academy, Strömmstad, Sweden
Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Melbourne, Australia
Department of Medicine, Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden
Center for Obesity, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

Corresponding Author:
Marie Löf, PhD
Department of Health Medicine and Caring Sciences
Linköping University
Linköping, 581 83 Sweden
Phone: 46 73 656 9338
Email: marie.lof@liu.se

Abstract

Background: Active transportation (AT; ie, walking and cycling as a mode for transportation) has been associated with decreased morbidity and mortality; however, low-cost and scalable intervention programs are lacking.

Objective: The goal of the research was to determine the effectiveness of a 3-month behavior change program delivered via a mobile phone app to promote AT (TravelVu Plus) on time spent in moderate-to-vigorous physical activity (MVPA).

Methods: For this 2-arm parallel randomized controlled trial, we recruited a population-based sample of 254 adults from Stockholm County who were aged 20 to 65 years and had access to a smartphone. On completion of 1-week baseline measures, the 254 participants were randomized to either the control or intervention group (1:1 ratio). Both groups had access to the standard TravelVu app (Trivector AB) for monitoring their AT for 6 months. The intervention group also received a 3-month behavior change program to promote AT (TravelVu Plus app). Assessors of outcomes were blinded to group allocation. Outcomes were objectively measured MVPA at 3 (primary) and 6 months. Secondary outcomes were AT, attitudes toward AT, and health-related quality of life at 3 and 6 months.

Results: No effect on MVPA was observed after 3 months ($P=.29$); however, at 6 months the intervention group had a greater improvement in MVPA than the controls (6.05 minutes per day [95% CI 0.36 to 11.74; $P=.04$]). A Bayesian analyses showed that there was a 98% probability that the intervention had any effect at 6 months, and a 63% probability that this effect was $>$5 minute MVPA per day.

Conclusions: No effect on MVPA immediately after the intervention period (at 3 months) was observed; however, there was a delayed effect on MVPA (6 minutes per day) at 6 months, which corresponds to approximately 30% of the weekly MVPA recommendation. Our findings suggest that a behavior change program promoting AT delivered via an app may have a relevant effect on PA.

Trial Registration: ClinicalTrials.gov NCT03086837; https://clinicaltrials.gov/ct2/show/NCT03086837
Introduction

Physical inactivity (little or no physical activity) is a major risk factor for noncommunicable diseases, including cardiovascular disease, type 2 diabetes, and premature death [1] and contributes to high health care costs in both high- and low-income countries [2]. Low-cost, scalable interventions aimed at increasing habitual physical activity (PA) at the population level are warranted. Active transportation (AT; ie, cycling and walking as mode of transportation) represents a key target since AT is easily accessible and enables regular PA on a daily basis [3]. A recent meta-analysis (23 prospective studies, n=531,333) concluded that AT was associated with decreased mortality and lower risks of cardiovascular disease and diabetes [4]. Also, randomized controlled trials (RCTs) support that AT by bicycle can improve health markers such as insulin sensitivity and cardiopulmonary fitness [5-7]. However, to date, behavioral interventions targeting AT to increase daily PA in adults are few, and trials have been of mixed quality, with considerable variation in sample characteristics, study duration, and outcomes [8-12]. Furthermore, population-based RCTs assessing the effects of AT on PA in healthy adults using objective measures for both AT and PA are lacking.

Mobile health (mHealth) interventions are increasingly being used to promote healthy lifestyle behaviors [13,14] and include the use of mobile apps [15], which offer potential to be delivered at scale. To the best of our knowledge, only one app has focused on promotion of AT. In that study, Bopp et al [12,16] evaluated whether a campaign with self-monitoring of AT (via an app) together with social media and marketing components could increase AT among students and employees at a university campus. Results showed an increase in the number of self-reported active trips by students. However, due to the multicomponent nature of the intervention [12], it was not possible to evaluate the effect of the app component alone. Also, it was a single group study and only included self-reported travel. Thus, further well-conducted RCTs evaluating whether an app can promote AT are warranted.

This paper reports the results of the Smart City Active Mobile Phone Intervention (SCAMPI) trial. The aim of this trial was to determine the effectiveness of a 3-month mobile phone–based behavior change program promoting AT on moderate-to-vigorous physical activity (MVPA) in Swedish adults [17]. The primary outcome was MVPA at 3 months, while secondary outcomes included MVPA at 6 months as well as time spent in AT, perceptions about AT, and health-related quality of life at 3 and 6 months.

Methods

Study Design

A 2-arm parallel design RCT was conducted between September 2017 and September 2018. The study was approved by the regional research ethics committee in Stockholm (January 11, 2017: 2016/2403-31 and June 30, 2017: 2017/1373-32) and registered at ClinicalTrials.gov [NCT03086837]. Details on the development of the app and design of the SCAMPI trial are published elsewhere [17]. The study is reported according to the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and onLine Telehealth (CONSORT-EHEALTH) statement [18].

Participants and Procedures

Participants were recruited from a random sample of 4995 adults provided by Statistics Sweden. Two waves of invitation letters were sent out (September 2017 [n=2000] and January 2018 [n=2995]) in order to capture different seasons and weather. Participants were eligible if they were aged 20 to 65 years, understood written Swedish, lived in the county of Stockholm, and had access to a smartphone compatible with the app. Exclusion criteria was not being able to perform MVPA. People who wanted to participate in the study signed up at the study website or sent an email or a letter to the research group. After providing informed consent (electronically signed), a web-based questionnaire was administrated to collect self-reported sociodemographic variables (eg, age, sex, country of birth, educational attainment), height, weight, attitude toward AT, neighborhood walkability, and health-related quality of life. Thereafter, baseline measures of PA (using the wGT3X-BT accelerometer [ActiGraph LLC]) as well as AT (measured through the TravelVu app [Trivektor AB], standard version) were assessed simultaneously during 7 consecutive days. The details of the TravelVu app have been provided in the study protocol [17]. Briefly, the app passively collected data using GPS coordinates and presented total and mode-specific travel minutes per day (for walking, cycling, car, train, ferry, or bus). At the end of each day during these 7 days, participants were asked to review and, if necessary, manually revise travel and locations in the app. After making the needed revisions to their travel routes, participants were instructed to approve these days by marking them as valid. All outcome measures were repeated postintervention at 3 months (primary time point) and 6 months postrandomization. An overview of the study design is provided in Figure 1.
Sample Size, Randomization, and Blinding
A total of 250 participants (125 per group) was estimated to provide 80% power ($\alpha=.05$) to detect a 10-minute difference in MVPA per day assuming a standard deviation of 25 minutes [19] and loss to follow-up of about 20%. On completion of baseline measures, participants were randomized to either the control or intervention group (1:1 ratio) using a computer-generated random allocation sequence list generated by the study statistician [17]. Allocation concealment was ascertained through opaque envelopes (ML). CA and ES enrolled participants. Assessors of outcomes were blinded; however, the group allocation was not blinded to the participants, who received an email after randomization.

Control Group
After baseline measures, participants in the control condition were encouraged to continue with their normal travel routines during the 6-month study period. During this time they were able to continue to monitor their daily travels using the standard TravelVu app (without a behavior change program) if they chose to. At 3 and 6 months, participants in the control group (as well as in the intervention group) were asked to monitor their travel behavior in the app during the same seven days as they wore the accelerometer for follow-up outcome assessment.

Intervention
In addition to the standard TravelVu app, intervention participants received a 3-month behavior change program (TravelVu Plus), aimed at increasing PA through AT. The program was delivered as extra features to the standard TravelVu app. The development of the TravelVu Plus program and its features are described in more detail in our study protocol [17]. Briefly, it was anchored in social cognitive theory [20] as well as social ecological principles [21] and included features such as a goal-setting function, messages (sent as push notifications), and feedback. In-app features encouraged participants to set new AT goals on a weekly basis. Feedback on participants’ AT and progress toward the set weekly goal were provided in graphical form throughout the week [17]. At the end of the week, push notifications were used to provide textual feedback on AT performance (ie, feedback on behavior) if PA goals were reached (ie, feedback on outcome of behavior) or to encourage modifying goal according to achievement (ie, review behavior goal). Furthermore, information on all achieved weekly goals thus far was provided graphically. Push notifications were also sent with general information on AT and its health and climate benefits as well as practical tips and behavior change strategies (Multimedia Appendix 1). After 3 months, the enhanced features were disabled and the intervention group had access to the standard version (ie, the TravelVu app) for the remaining study period (3 to 6 months after baseline) and could continue to monitor their AT. This enabled us to assess to what extent they used the self-monitoring of AT during the follow-up period and whether this was different from the control group.

Outcomes
PA was measured objectively using the wGT3X-BT triaxial accelerometer (ActiGraph LLC), which was worn on the hip over seven consecutive 24-hour periods. Raw acceleration data (at 90 Hz) were uploaded and processed using the ActiLife software version 6.13.3 (ActiGraph LLC) into filtered sum of vector magnitudes (VM). Nonwear time was detected and excluded using a Troiano algorithm [22]. A day was categorized as valid if wear time $\geq$600 minutes [23]. For each participant, time spent in light PA (VM 201-2690 counts per minute), moderate PA (MPA; VM 2691-6166 counts per minute),
vigorous PA (VM: ≥6167 counts per minute), and MVPA (VM ≥2691 counts per minute) were calculated using recommended cutoffs [23] by Sasaki et al [24], while time spent sedentary (VM 0-200 counts per minute) was derived by applying cutoffs by Aguilar-Farias et al [25].

To complement the accelerometer data, we also evaluated the number of minutes spent cycling and walking for transport assessed using the TravelVu app during seven consecutive 24-hour periods as a secondary outcome. Days marked as valid were used to assess mode and duration (minutes per day) of AT (cycling and walking). Days with unreasonably high levels of AT for the Stockholm area (>4 hours) were not included. Attitudes toward AT were assessed using the psychosocial items in section B of the validated Transport and Physical Activity Questionnaire [26]. Mean values for each AT mode (walking and cycling) were calculated for each participant. The RAND-36 was used to assess health-related quality of life [27,28]. The general health domain was analyzed in this study.

Other Measures
Perceived neighborhood walkability was assessed at baseline using the Neighborhood Environment Walkability Scale questionnaire [29,30]. To calculate a perceived walkability index for each participant, we summed the z scores for residential density, street connectivity, and land use mix as described previously [19]. Finally, app engagement was measured as the number of registered days in the app (ie, days that the participant had reviewed and approved as valid regarding their travel behavior that day; intervention and control group) as well as the number of goals set and achieved (intervention group only).

Statistical Analysis
All statistical analyses were conducted in accordance with the study protocol [17] and followed intention-to-treat principles. Linear mixed models (random intercept) were used to contrast differences in primary (MVPA) and secondary outcomes (AT, RAND-36 general health, attitudes toward using AT) between the intervention and control group. Outcomes were regressed against group allocation and included a group × time interaction term to incorporate repeated measures (0, 3, and 6 months).

Three-way interaction analyses were performed to assess if the following characteristics at baseline moderated the intervention effect on the primary outcome: sex, age, educational attainment, BMI, foreign background, season of randomization, perceived walkability index, attitude toward AT, or general health. We also examined whether the effect on MVPA was associated with engagement with the app (number of registered days as well as number of goals set and achieved in the app).

Due to relatively few missing values in the outcome measures (30/252, 11.9%, and 34/252, 13.5%, at 3- and 6-month follow-ups, respectively) and since we could not rule out the possibility that data were missing at random, we followed the recommendations to report completers only as the primary analyses [31-33]. A sensitivity analysis where missing data for the primary outcome at the 3- and 6-month follow-ups were imputed using multiple imputation with chained equations [33] (predictive mean matching, with 500 imputations and 30 iterations) was also conducted. Deviations from the missing completely at random assumption were evaluated through attrition analysis where baseline characteristics for completers and noncompleters were compared.

We also conducted the following analyses that were added to the statistical analysis plan before data analysis but not reported in the protocol [17]. First, since AT mainly corresponds to MPA, the largest component of MVPA, we also used linear mixed models to contrast MPA between the two groups. Furthermore, recent data indicate that even light PA may reduce premature mortality [34-36], and since we cannot exclude that some AT would be light PA, we also ran models with light PA as outcome. Additionally, we explored whether accelerometer wear time influenced our results; however, our estimates remained similar after adjustment (results not shown). Finally, in exploratory analyses, Bayesian inference for the linear mixed models was employed to calculate the probability that the intervention had an effect on MVPA [37]. These Bayesian analyses provided a more robust view of the data collected in the trial due to the following reasons: (1) P values and confidence intervals are not well defined in linear mixed models [38] and should therefore only be taken as approximate and (2) null hypothesis testing can be sensitive to individual data points [39]. Uniform priors were used for all parameters in the Bayesian analyses.

Statistical analyses were performed with a significance level of .05 using R version 3.6.1, and Bayesian inference was done using the probabilistic programming language Stan (RStan version 2.19.1, both R Foundation for Statistical Computing and SPSS Statistics version 24 (IBM Corporation).

Results
Participants
Figure 2 presents the flow of the participants. In total, 473 out of 4995 responded to the invitation letter and 254 completed baseline measures and were randomized (1:1). All accelerometer files went through an overall review to ascertain sufficient data recordings before randomization; however, when the accelerometer data were processed in detail after the study was completed, it was discovered that two participants in the control group did not fulfill the wear time criteria (≥600 minutes per day). Therefore, they were excluded from the analyses, and thus the final sample was n=252. No major differences were found between the 254 participants and the invited population-based sample regarding area of residence (city center or countryside) and age. However, participants were more often women, born in Sweden, and had a university degree compared with nonparticipants (Multimedia Appendix 2). There were no differences between the intervention and control group with respect to baseline characteristics (Table 1).
Figure 2. Flowchart of the Smart City Active Mobile Phone Intervention trial.

Received inquiry letter about the study (N=4995)

Responded to letter and assessed for eligibility (n=473)

Excluded (n=219)
- Ineligible (n=10)
- Declined participation due to stress (n=16) or not interested (n=149)
- Did not complete baseline measures (n=44)

Baseline

Randomization (1:1)
(n=254)

Allocated to intervention (n=127)
TRavelVU app with a 3-month behaviour change programme

Allocated to control (n=127)
TRavelVU app (no behaviour change programme)

Excluded at baseline (n=2)
- Invalid accelerometer data (n=2)

3-month follow-up

Lost to follow-up (n=15)
- Declined further participation due to:
  ⇒ loss of family member (n=2)
  ⇒ lack of time/stress (n=3)
  ⇒ did not want to use the app (n=2)
  ⇒ no reason (n=5)
- No contact (n=3)

Missing (n=2)
- Accelerometer lost in mail (n=1)
- Insufficient accelerometer data (n=1)

6-month follow-up

Missing (n=5)
- Declined further participation due to:
  ⇒ less of family member (n=1)
  ⇒ disease (n=1)
- No contact (n=3)

Missing (n=3)
- Declined further participation due to:
  ⇒ lack of time/stress (n=1)
  ⇒ disease (n=1)
- No contact (n=1)

Analyses

Included in the analysis at:
- Baseline (n=127)
- 3-months (n=110)
- 6-months (n=107)

Included in the analysis at:
- Baseline (n=125)
- 3-months (n=112)
- 6-months (n=111)
Table 1. Baseline characteristics of the participants in the Smart City Active Mobile Phone Intervention trial.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (n=125)</th>
<th>Intervention (n=127)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, mean (SD)</td>
<td>46.2 (11.0)</td>
<td>46.5 (11.0)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>66 (52.8)</td>
<td>78 (61.4)</td>
</tr>
<tr>
<td>Male</td>
<td>59 (47.2)</td>
<td>48 (37.8)</td>
</tr>
<tr>
<td>Other</td>
<td>0 (0)</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>2 (1.6)</td>
<td>3 (2.4)</td>
</tr>
<tr>
<td>Secondary</td>
<td>42 (33.6)</td>
<td>44 (34.6)</td>
</tr>
<tr>
<td>Tertiary</td>
<td>81 (64.8)</td>
<td>80 (63.0)</td>
</tr>
<tr>
<td>Body mass index (kg/m²), mean (SD)</td>
<td>24.7 (3.1)</td>
<td>24.9 (4.0)</td>
</tr>
<tr>
<td>Sedentary ab (min/d), mean (SD)</td>
<td>485.4 (69.7)</td>
<td>477.9 (83.4)</td>
</tr>
<tr>
<td>Light activity ab (min/d), median (IQR)</td>
<td>310.8 (69.7)</td>
<td>320.6 (83.9)</td>
</tr>
<tr>
<td>Moderate to vigorous activity ab (min/d), mean (SD)</td>
<td>60.3 (26.0)</td>
<td>59.7 (27.6)</td>
</tr>
<tr>
<td>Moderate activity ab (min/d), mean (SD)</td>
<td>52.0 (22.0)</td>
<td>50.9 (23.2)</td>
</tr>
<tr>
<td>Vigorous activity ab (min/d), mean (SD)</td>
<td>8.2 (12.2)</td>
<td>8.8 (12.3)</td>
</tr>
<tr>
<td><strong>Active transportation c (min/d), mean (SD)</strong></td>
<td>56.8 (26.5)</td>
<td>58.1 (28.1)</td>
</tr>
<tr>
<td>Walking (min/d), mean (SD)</td>
<td>50.5 (25.8)</td>
<td>54.2 (27.9)</td>
</tr>
<tr>
<td>Cycling (min/d), mean (SD)</td>
<td>6.3 (16.1)</td>
<td>3.9 (9.2)</td>
</tr>
<tr>
<td>General health d, mean (SD)</td>
<td>75.3 (17.7)</td>
<td>73.8 (17.1)</td>
</tr>
<tr>
<td>Perceived walkability e, mean (SD)</td>
<td>-0.2 (2.3)</td>
<td>0.2 (2.4)</td>
</tr>
<tr>
<td><strong>Attitude toward f, mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking</td>
<td>3.9 (0.6)</td>
<td>4.0 (0.5)</td>
</tr>
<tr>
<td>Cycling</td>
<td>3.5 (0.8)</td>
<td>3.5 (0.8)</td>
</tr>
<tr>
<td>Counts per minute, mean (SD)a</td>
<td>389 (137.8)</td>
<td>390 (137.7)</td>
</tr>
</tbody>
</table>

abMeasured by accelerometer.
Wear time (days) for the accelerometer was 6.4 (SD 1.2) days (intervention) and 6.5 (SD 1.1) days (control). The corresponding wear time in minutes per day was 853 (SD 67) minutes (intervention) and 857 (SD 73) minutes (control).
cMeasured by TravelVu smartphone app. Control (n=124) and intervention (n=126) due to missing data.
dMeasured by RAND-36.
eMeasured by means of Neighborhood Environment Walkability Scale.
fMeasured by means of Transport and Physical Activity Questionnaire.

Effectiveness of the Intervention

Results showed no statistically significant difference between groups on the primary outcome at 3 months (P=.29); however, at 6 months, the intervention group had a greater improvement in MVPA than the control group (6.05 minutes per day; 95% CI 0.36 to 11.74; P=.04). As shown in Figure 3, the difference in MVPA at 6 months was driven predominantly by changes in MPA (difference 7.21 minutes per day; 95% CI 1.95 to 12.47; P=.007). Sensitivity analysis (imputed data) revealed comparable results (Multimedia Appendix 3). There was an interaction between sex and the 6-month MVPA (group × time × gender coefficient estimate 14.7 minutes per day; 95% CI 3.2 to 26.1; P=.01), indicating the intervention was more effective in men than in women. There was no interaction effect for any of the other investigated covariates (ie, age, educational attainment, BMI, foreign background, season of randomization, perceived walkability index, attitude toward AT, or general health [results not shown]). AT (minutes per day) was statistically significantly associated with accelerometer MVPA (minutes per day; r=0.5; P<.001) at baseline, and this association remained similar at the two follow-ups. Pre-post differences of MVPA did not differ by app engagement (number of registered days in the app, goals set or achieved; results not shown). Table 2 presents the corresponding results for the secondary outcomes. No differences in the change of other secondary outcomes
between follow-up and baseline were found between the groups. Correspondingly, no difference in the change of AT walking or AT cycling for transport was observed (results not shown). Finally, no statistically significant difference was found when contrasting light PA between groups.

Figure 4 reports the results from the Bayesian analyses for MVPA. The probability that the intervention group improved MVPA more than the control group (ie, had any effect on MVPA) was 84.8% at 3 months and 97.8% at 6 months. Furthermore, the probability that this improvement was more than 5 minutes per day was 63.3% at 6 months.

Figure 3. Intervention effect on moderate-to-vigorous physical activity and moderate physical activity at 3 and 6 months.
## Table 2. The intervention effect on the secondary outcomes at 3 and 6 months.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Sample, mean (SD)</th>
<th>Group x time interaction&lt;sup&gt;a&lt;/sup&gt;</th>
<th>95% CI&lt;sup&gt;b&lt;/sup&gt;</th>
<th>P value&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Intervention</td>
<td>Group x time</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Active transportation</strong>&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months&lt;sup&gt;d&lt;/sup&gt;</td>
<td>58.0 (30.5)</td>
<td>58.6 (29.7)</td>
<td>0.96</td>
<td>-6.91 to 8.80</td>
</tr>
<tr>
<td>6 months&lt;sup&gt;e&lt;/sup&gt;</td>
<td>58.3 (29.0)</td>
<td>60.5 (32.4)</td>
<td>2.02</td>
<td>-6.39 to 10.4</td>
</tr>
<tr>
<td><strong>Attitude toward walking</strong>&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months&lt;sup&gt;g&lt;/sup&gt;</td>
<td>3.9 (0.6)</td>
<td>4.0 (0.6)</td>
<td>0.02</td>
<td>-0.11 to 0.14</td>
</tr>
<tr>
<td>6 months&lt;sup&gt;h&lt;/sup&gt;</td>
<td>4.0 (0.6)</td>
<td>4.0 (0.6)</td>
<td>-0.12</td>
<td>-0.25 to 0.01</td>
</tr>
<tr>
<td><strong>Attitude toward cycling</strong>&lt;sup&gt;f&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months&lt;sup&gt;i&lt;/sup&gt;</td>
<td>3.6 (0.9)</td>
<td>3.5 (1.0)</td>
<td>-0.1</td>
<td>-0.25 to 0.06</td>
</tr>
<tr>
<td>6 months&lt;sup&gt;j&lt;/sup&gt;</td>
<td>3.7 (0.9)</td>
<td>3.5 (0.9)</td>
<td>-0.1</td>
<td>-0.26 to 0.05</td>
</tr>
<tr>
<td><strong>General health</strong>&lt;sup&gt;k&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 months&lt;sup&gt;l&lt;/sup&gt;</td>
<td>76.9 (18.5)</td>
<td>76.5 (16.7)</td>
<td>1.52</td>
<td>-1.69 to 4.74</td>
</tr>
<tr>
<td>6 months&lt;sup&gt;m&lt;/sup&gt;</td>
<td>75.3 (18.5)</td>
<td>77.2 (17.1)</td>
<td>3.17</td>
<td>-0.11 to 6.44</td>
</tr>
</tbody>
</table>

<sup>a</sup>Fixed effect coefficient estimate from linear mixed model (random intercept).

<sup>b</sup>Given by 2-sided Wald test.

<sup>c</sup>Measured by TravelVu (smartphone app).

<sup>d</sup>Control (n=102); intervention (n=106).

<sup>e</sup>Control (n=93); intervention (n=80).

<sup>f</sup>Measured by means of Transport and Physical Activity Questionnaire.

<sup>g</sup>Control (n=110); intervention (n=102).

<sup>h</sup>Control (n=107); intervention (n=95).

<sup>i</sup>Control (n=110); intervention (n=102).

<sup>j</sup>Control (n=107); intervention (n=95).

<sup>k</sup>Measured by RAND-36.

<sup>l</sup>Control (n=110); intervention (n=102).

<sup>m</sup>Control (n=107); intervention (n=95).
Figure 4. Bayesian analysis of the intervention effect on moderate-to-vigorous physical activity at 3 and 6 months.

App Engagement

Objective measures of engagement with the self-monitoring feature of AT in the standard TravelVu app for the two groups are shown in Table 3. For the intervention group, during the 3-month intervention period, 60.6% of participants (77/127) registered 57 days or more out of the 84 days in total, indicating high engagement. Also, as shown in Table 4, the goal setting function in the TravelVu Plus app was relatively well used with 58.2% of participants (74/127) in the intervention group setting weekly goals for 5 weeks or more and 46.4% (59/127) achieving those goals. App engagement decreased after the intervention period when participants only had access to the standard version (TravelVu); however, many participants in the intervention group continued to use the self-monitoring feature for the subsequent 3 months (Table 3).

In the control group (TravelVu), as seen in Table 3, 65.6% (82/125) continued to register days with AT in the app beyond the baseline assessment; however, the number of days (mean 39 [SD 35]) was fewer than for the intervention group (mean 53 [SD 35]; \( P = .01 \)) during the first 3 months. During months 3 to 6 after baseline, the number of days were comparable (control: mean 34 [SD 34] days vs intervention: mean 30 [SD 32] days; \( P = .37 \)).

Table 3. Objective measures of engagement with the self-monitoring feature of AT in the standard version of the app (TravelVu) in the intervention group (n=127) and control group (n=125).

<table>
<thead>
<tr>
<th>Number of registered days in the appa,b,c</th>
<th>Intervention group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 to 3 months, n (%)</td>
<td>3 to 6 months after intervention, n (%)</td>
</tr>
<tr>
<td>57-84</td>
<td>77 (60.6)</td>
<td>29 (22.8)</td>
</tr>
<tr>
<td>29-56</td>
<td>16 (12.6)</td>
<td>13 (10.2)</td>
</tr>
<tr>
<td>7-28</td>
<td>10 (7.9)</td>
<td>43 (33.9)</td>
</tr>
<tr>
<td>0-6</td>
<td>24 (18.9)</td>
<td>42 (33.1)</td>
</tr>
</tbody>
</table>

\( a \) A registered day is defined as a day that was reviewed and approved by participant as valid data regarding their travel behavior that day (ie, number of minutes spent walking, cycling).

\( b \) The maximum total number of days was 84 days since results are reported for 3 months or 12 weeks (ie, 0-3 months [intervention period] or 3-6 months [3-6 months after the intervention]).

\( c \) These categories correspond to <1 week, 1-4 weeks, 4-8 weeks and 8-12 weeks.
Discussion

Principal Findings

This is the first RCT to determine the effect of a stand-alone behavior change program promoting AT delivered through a mobile phone app (ie, TravelVu Plus) on MVPA among adults. Access to the TravelVu Plus app did not lead to any statistically significant difference in MVPA at 3 months (primary outcome); however, relevant differences in MVPA were observed at 6 months. This difference was driven predominantly by an increase in MPA, which is reasonable since the intervention targeted cycling and walking. Bayesian analyses provided further support for an effect of the intervention.

Comparison With Prior Work

To the best of our knowledge, only one previous study evaluated the effectiveness of an app designed to solely promote AT [12]. In that study, an increase in self-reported AT to campus among 563 university students was reported [12]; however, it was not possible to evaluate the effectiveness of the app alone since it was part of a multicomponent intervention. In our study, there are several possible explanations for why there was an intervention effect at 6 months but not at 3 months. First, the intervention group may have needed a longer time period to achieve a behavior change (ie, increase their AT). Thus, even though the intervention group was less engaged in the app after 3 months when the behavior change features were disabled, it may be speculated that they had found strategies during the intervention and then applied those strategies to increase AT (detectable at 6 months). Second, in accordance with the predetermined study protocol, both groups had access to the TravelVu app in order to objectively assess AT during 7 days at baseline and the two follow-ups. Although the control group did not receive the 3-month enhanced app features, it is possible that using the standard app, anchored in self-monitoring [40], made them aware of their PA pattern and thus influenced them to engage in more PA in the beginning of the study period and thus findings for AT as outcome should be interpreted with caution. Our original preference was to passively record AT by mobile app for the assessment period only (it would be shut off automatically between the baseline and follow-up periods); however, this was not technically feasible at that time. Future studies should consider how to address these issues to optimize objectively assessed AT.

Strengths, Limitations, and Generalizability

Strengths of the SCAMPI trial include the RCT design, as well as the objectively measured primary outcome (MVPA). Previous studies on AT and PA are mostly observational, and systematic reviews have called for well-designed interventions targeting AT [8,11,46]. The included Bayesian analyses further strengthen the hypothesis that the TravelVu Plus app helped individuals increase their MVPA.

The main limitation of this study is that the primary outcome (MVPA) was assessed by accelerometry which does not capture cycling behavior; however, this choice was carefully considered when designing the study. Notwithstanding its limitations, an objective measure of PA was considered preferable to self-report as a secondary outcome. This approach relied on participants’ approval of trips taken and making corrections to these as necessary (ie, ensuring the trip was undertaken as indicated in the app). Notwithstanding, participants in the intervention group approved 71% of their travel trips (60 out of 84 days), indicating high use of this feature to self-monitor their AT. Furthermore, although the control group only had access to the standard version of the app, they used it for an average of 39 days, which supports that they also appreciated monitoring AT. Still, given the required level of interaction, only 65% of participants provided complete data on AT for baseline and the two follow-up assessments. This loss of data likely contributed to not being able to detect an effect on AT, and thus findings for AT as outcome should be interpreted with caution. Our original preference was to passively record AT by a mobile app for the assessment period only (it would be shut off automatically between the baseline and follow-up periods); however, this was not technically feasible at that time. Future studies should consider how to address these issues to optimize objectively assessed AT.

Table 4. Number of set and achieved goals in the TravelVu Plus app by the intervention group (n=127) during the 3-month intervention perioda.

<table>
<thead>
<tr>
<th>Number of weekly goals in the appb</th>
<th>Set goals</th>
<th>Achieved goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-12</td>
<td>53 (41.7)</td>
<td>21 (16.5)</td>
</tr>
<tr>
<td>5-8</td>
<td>21 (16.5)</td>
<td>38 (29.9)</td>
</tr>
<tr>
<td>1-4</td>
<td>27 (21.3)</td>
<td>36 (28.3)</td>
</tr>
<tr>
<td>0</td>
<td>26 (20.5)</td>
<td>32 (25.5)</td>
</tr>
</tbody>
</table>

aThe number of goals set and achieved provided in the table were extracted from the app (ie, objectively measured).
bThe maximum number of goals is 12 since the intervention was 3 months (ie, 12 weeks).

It is also relevant to compare our results with other app-based PA interventions. Systematic reviews and meta-analyses have concluded that mobile phone–based interventions may have small to moderate effect sizes when it comes to improving PA in free-living adults [41-44]. A recent study showed a mean increase in daily step count of between 226 and 319 steps associated with four different types of mobile phone interventions, which equated to around 5% of the mean daily step count for American adults (4700 steps per day) [45]. Thus, our findings are comparable to previous app-based PA interventions and extend available literature since we only targeted AT as a PA behavior.

To the best of our knowledge, this is the first trial that evaluated objectively assessed AT via GPS monitoring instead of using self-report as a secondary outcome. This approach relied on participants’ approval of trips taken and making corrections to these as necessary (ie, ensuring the trip was undertaken as indicated in the app). Notwithstanding, participants in the intervention group approved 71% of their travel trips (60 out of 84 days), indicating high use of this feature to self-monitor their AT. Furthermore, although the control group only had access to the standard version of the app, they used it for an average of 39 days, which supports that they also appreciated monitoring AT. Still, given the required level of interaction, only 65% of participants provided complete data on AT for baseline and the two follow-up assessments. This loss of data likely contributed to not being able to detect an effect on AT, and thus findings for AT as outcome should be interpreted with caution. Our original preference was to passively record AT by a mobile app for the assessment period only (it would be shut off automatically between the baseline and follow-up periods); however, this was not technically feasible at that time. Future studies should consider how to address these issues to optimize objectively assessed AT.
captured with accelerometry. Thus, it is reasonable to conclude that most likely our effects on MVPA are slightly underestimated due to the fact that the accelerometer could not capture cycling. Also, we cannot exclude that some AT could be light PA instead of MVPA; however, no difference was found when contrasting light PA between groups. Another important limitation to consider is seasonality. Sweden has a relatively cold climate with a winter period between December and February; however, the autumn and spring months can also be quite cold and snowy, which does not facilitate AT. As described above, we recruited in two waves (early autumn and early spring) in order to spread the intervention period out throughout different seasons, and we did not find any evidence that the effectiveness of the intervention differed upon which season participants entered the trial. Finally, other limitations of this trial include that automated feedback messages were used instead of personalized, that the review and correction of the automatically captured AT was considered time-consuming by some participants, and that the app consumed quite a lot of battery power (as indicated in the postintervention qualitative interviews, reported separately).

To minimize the risk for selection bias, recruitment was population-based with a sample randomly drawn by Statistics Sweden. Nevertheless, as is common in research, the willingness to participate was higher among women and well-educated people. However, the effect was not moderated by baseline educational attainment. Furthermore, both groups had on average 60 minutes MVPA at baseline. No comparable population-based data regarding MVPA exists in Sweden, and comparisons between studies are difficult due to different protocols and accelerometer cut points; however, previous studies in Swedish adult populations have shown approximately 35 minutes of MVPA per day [47-49]. Thus, although a random sample was drawn, we cannot exclude that people who signed up had a relatively physically active lifestyle, and the potential of the TravelVU Plus app to promote AT in more sedentary populations should be explored. Also, we cannot generalize our results to people who recently migrated to Sweden since the app is currently only available in Swedish. Finally, it is relevant to note that Stockholm is a city amenable to AT (walking and cycling); however, many cities and suburbs across the world are not, which may limit generalizability to other types of cities as well as limit the reach of interventions targeting only AT. Thus, future studies should evaluate the potential of apps to promote AT similar to the TravelVU Plus app in other types of cities and contexts.

Implications and Future Research

The SCAMPI trial provides several important findings and insights for future interventions targeting AT. First, objective measures showed that the TravelVU Plus app was used frequently for 3 months, indicating its potential for promoting AT. This finding was supported by the in-depth interviews with a subsample (reported separately). Furthermore, this qualitative data suggested the need for improvements such as personalized messages and an improved function for registrations, preferably automated if possible, for correction of travel trips in the app. These improvements may enhance the effectiveness of the app. Second, we were able to observe a 6-minute difference in MVPA in the intervention group compared with the control group at 6 months. The observed effect is modest; however, considering that 6 minutes per day adds up to 42 minutes per week which corresponds to nearly 30% of the recommendation (150 minutes MVPA per week) [50] and it was achieved through a low-cost and scalable intervention, this finding is important. In this context it is also important to highlight that, due to the 24-hour continuum, adding time in AT is a reduction of time spent in other activities—in this case most likely a reduction in sedentary time (replacing motor transport) which also has positive health implications. Previous data indicate that Swedish adults spend as much as 60% of their total time as sedentary [47] and also that very small substitutions of sedentary time for MPA have shown decreased risks for metabolic syndrome [51,52]. Also, it may be speculated that the observed effect might have been larger if the control group did not have access to the app for the measurement of AT. Thus, it is reasonable to conclude that our findings together with the high app engagement motivates further investigation of the potential of the TravelVu Plus app to promote AT.

Conclusions

We observed no effect on MVPA immediately after the intervention period (at 3 months); however, there was a delayed effect on MVPA (6 minutes per day) at 6 months, which corresponds to almost 30% of the weekly MVPA recommendation. Our findings coupled with the high engagement in the app suggest that a behavior change program promoting AT delivered through an app may have a relevant effect on PA, motivating further research on mHealth AT interventions.

Acknowledgments

This study was funded by Forte (2016-00138) and Karolinska Institutet (2018-01730). They had no role in study design, conduct, or reporting. We thank all the participants in the trial as well as Emeli Adell and Leif Linse at Trivector AB for technical support with the TravelVU app and with the development of the behavior change program (TravelVu Plus app). We also thank Cristina Cadenas-Sánchez and Jairo Migueles for support with accelerometer data processing and André Lauber, who helped us set up the online questionnaires.

Authors’ Contributions

ML is the principal investigator of this RCT and conceptualized and designed the study in collaboration with RM and UE. AE, CA, and ES collected the data. ML and AE designed the intervention with contribution from CA, UE, and AD. PB contributed
to accelerometer data processing. MB conducted the statistical analyses. PH, YLT, and CDN contributed to the interpretation of results. AE and ML drafted the manuscript, which was reviewed and approved by all coauthors.

**Conflicts of Interest**
The authors have no conflict of interest. MB owns a private company (Alexit AB), which develops and disseminates eHealth apps to health organizations and professionals in both the private and public sector; however, Alexit AB was not involved in any part of this study.

**Multimedia Appendix 1**
Screenshots from the app showing examples of messages, weekly summary of travel and number of achieved weekly goals.

**Multimedia Appendix 2**
Supplementary Table 1.

**Multimedia Appendix 3**
Supplementary Table 2.

**Multimedia Appendix 4**
CONSORT-EHEALTH checklist (V 1.6.1).

**References**


36. Saint-Maurice PF, Troiano RP, Berrigan D, Kraus WE, Matthews CE. Volume of light versus moderate-to-vigorous physical activity: similar benefits for all-cause mortality? J Am Heart Assoc 2018 Apr 02;7(7) [FREE Full text] [doi: 10.1161/JAHA.118.008815] [Medline: 29610219]


**Abbreviations**

AT: active transportation
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

**mHealth**: mobile health

**MPA**: moderate physical activity

**MVPA**: moderate-to-vigorous physical activity

**PA**: physical activity

**RCT**: randomized controlled trial

**SCAMPI**: Smart City Active Mobile Phone Intervention

**VM**: vector magnitude

---

Edited by G Eysenbach; submitted 03.03.20; peer-reviewed by J Rossen, D Berrigan; comments to author 16.03.20; revised version received 17.04.20; accepted 20.04.20; published 08.06.20.

Please cite as:
Effectiveness of a 3-Month Mobile Phone–Based Behavior Change Program on Active Transportation and Physical Activity in Adults: Randomized Controlled Trial
JMIR Mhealth Uhealth 2020;8(6):e18531
URL: https://mhealth.jmir.org/2020/6/e18531
doi:10.2196/18531
PMID:32510462

©Anna Ek, Christina Alexandrou, Emmie Söderström, Patrick Bergman, Christine Delisle Nyström, Artur Direito, Ulf Eriksson, Pontus Henriksson, Ralph Maddison, Ylva Trolle Lagerros, Marcus Bendtsen, Marie Löf. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 08.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Health Gain, Cost Impacts, and Cost-Effectiveness of a Mass Media Campaign to Promote Smartphone Apps for Physical Activity: Modeling Study

Anja Mizdrak¹, BA, DPhil; Kendra Telfer¹, Hons; Artur Direito², PhD; Linda J Cobiac³, PhD; Tony Blakely¹,4, MBChB, NZPHM, PhD; Christine L Cleghorn¹, BSc, MSc, PhD; Nick Wilson¹, MPH

¹Department of Public Health, University of Otago Wellington, Wellington, New Zealand
²Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
³Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom
⁴Centre for Epidemiology and Biostatistics, University of Melbourne, Melbourne, Australia

Corresponding Author:
Anja Mizdrak, BA, DPhil
Department of Public Health
University of Otago Wellington
Department of Public Health, 23 Mein Street
Newtown
Wellington, 6021
New Zealand
Phone: 64 49186192
Email: anja.mizdrak@otago.ac.nz

Abstract

Background: Physical activity smartphone apps are a promising strategy to increase population physical activity, but it is unclear whether government mass media campaigns to promote these apps would be a cost-effective use of public funds.

Objective: We aimed to estimate the health impacts, costs, and cost-effectiveness of a one-off national mass media campaign to promote the use of physical activity apps.

Methods: We used an established multistate life table model to estimate the lifetime health gains (in quality-adjusted life years [QALYs]) that would accrue if New Zealand adults were exposed to a one-off national mass media campaign to promote physical activity app use, with a 1-year impact on physical activity, compared to business-as-usual. A health-system perspective was used to assess cost-effectiveness, and a 3% discount rate was applied to future health gains and health system costs.

Results: The modeled intervention resulted in 28 QALYs (95% uncertainty interval [UI] 8-72) gained at a cost of NZ $81,000/QALY (2018 US $59,500; 95% UI 17,000-345,000), over the remaining life course of the 2011 New Zealand population. The intervention had a low probability (20%) of being cost-effective at a cost-effectiveness threshold of NZ $45,000 (US $32,900) per QALY. The health impact and cost-effectiveness of the intervention were highly sensitive to assumptions around the maintenance of physical activity behaviors beyond the duration of the intervention.

Conclusions: A mass media campaign to promote smartphone apps for physical activity is unlikely to generate much health gain or be cost-effective at the population level. Other investments to promote physical activity, particularly those that result in sustained behavior change, are likely to have greater health impacts.

(JMIR Mhealth Uhealth 2020;8(6):e18014) doi:10.2196/18014

KEYWORDS
physical activity; mHealth; mobile health; smartphone apps; modeling; mass media campaigns

Introduction

Insufficient physical activity is associated with an increased risk of cardiovascular diseases, cancers, and poor mental health [1-3]. International recommendations state that adults should aim to accumulate at least 150 minutes of moderate-to-vigorous physical activity (MVPA) throughout the week [1,4]. Prevalence of insufficient physical activity is high in many countries: 40%

http://mhealth.jmir.org/2020/6/e18014/
in the United States, 34% in India, 47% in Brazil, and 42% in New Zealand [5]. Strategies to increase physical activity at the population level are needed, and the promotion of smartphone apps for physical activity is one promising avenue for intervention.

The rise of physical activity smartphone apps provides an opportunity to deliver interventions that have wide reach and a range of technology-enhanced features (eg, accelerometers, tailored feedback, and reminders) [6]. Evaluations of the effectiveness of physical activity apps have shown they can be effective at increasing physical activity levels [7-9]. However, there is high variability in the quality and effectiveness of the thousands of physical activity apps that are currently available [6,10]. Encouraging the use of high-quality apps provides a potential opportunity to increase population-level physical activity owing to the large potential reach and low cost of apps. Additionally, there is growing evidence of the cost-effectiveness of mobile health interventions as a whole [11].

Recent attempts have been made to improve public awareness around the quality and effectiveness of different health apps. Several government agencies around the world now provide app ratings or recommendations on their websites [12-16], but the levels of public engagement have not been publicly documented. Mass media campaigns provide a potential avenue to promote the use of high-quality physical activity apps and, thereby, result in increases in physical activity levels. A recent review suggests that mass media campaigns can be effective, but evidence on the cost-effectiveness is largely limited to tobacco control [17]. Our previous research has assessed the potential of mass media campaigns that promote smartphone apps: a mass media campaign promoting smoking cessation apps is likely to be cost-saving [18], while a mass media campaign for weight loss apps may or may not be cost-effective owing to wide uncertainty around intervention impacts [19]. Although the short-term effectiveness of physical activity apps has been assessed [7,8], it is unknown whether promoting physical activity apps through mass media would be effective or cost-effective. Similarly, we do not know how impacts of mass media campaigns to promote physical activity apps may compare to other public health interventions.

To fill this gap, this study assessed the health impacts, costs, and cost-effectiveness of a mass media campaign to promote high quality smartphone apps for physical activity in a high-income country setting (New Zealand) using a multistate life table modeling approach parameterized with age, sex, and ethnicity specific data consistent with previous work [18,19].

**Methods**

**Overview**

We used an established proportional multistate life table model to estimate the health impact of a mass media campaign to promote the use of physical activity smartphone apps [20,21]. The model simulates the entire New Zealand population, alive in 2011, out until death under both business-as-usual (BAU) and the modeled intervention. Health gain was measured in quality-adjusted life years (QALYs)—a summary measure of population health that captures both morbidity and mortality impacts to be considered simultaneously [22]. For costs, we used a health-system perspective, and the outputs were the difference in total health system costs (the net sum of intervention costs and downstream cost offsets due to altered disease rates) between BAU and the modeled intervention.

A 3% discount rate was applied to both health gains and health system costs in accordance with prior New Zealand research. Results for 0% and 6% discount rates are presented as scenario analyses. Full details of the model are published elsewhere [20,21].

**Intervention Specification**

We modeled a one-off mass media campaign according to the intervention pathway displayed in Figure 1. The population eligible for the intervention included all New Zealand adults 15-79 years of age—the population for which physical activity data were available. Of the population eligible for the intervention, we estimated the proportion of the population that would experience increased physical activity based on likely awareness of the mass media campaign, app download rates, and app use. We defined app use as use for at least 7 days following the initial download of the app to ensure consistency between the modeled intervention pathway and available evidence. Increases in physical activity associated with app use were estimated from a recent systematic review with a meta-analysis of the effectiveness of physical activity apps [8]. Increases in app use wane over time [23], with no evidence of maintained effect beyond 1 year [8]. As our model projects health gains in 1-year time steps, we estimated an average adherence to physical activity apps across the year in which the intervention was implemented [24]. Sources of parameter values are detailed in Table 1, and further detail on parameter selection is available in a related technical report [24].
Figure 1. Flow chart of intervention conceptualization. Italicized text represents the percentage of the eligible population exposed to each step in the intervention pathway. NZ: New Zealand.

![Flow chart of intervention conceptualization](image)

Table 1. Intervention parameters and uncertainty distributions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Distribution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult NZ(^a) population aware of mass media campaign, % (UI)(^b)</td>
<td>77.9 (70-83)</td>
<td>Beta</td>
<td>Based on awareness of previous health-related mass media campaign in NZ (Health Promotion Agency [25])</td>
</tr>
<tr>
<td>Adult NZ population who downloaded a physical activity app, % (UI)</td>
<td>31 (21-41)</td>
<td>Beta</td>
<td>Estimated based on the proportion of survey respondents who had downloaded a physical activity app to track behavior (Krebs and Duncan [26])</td>
</tr>
<tr>
<td>Adult NZ population who used the physical activity app, % (UI)</td>
<td>16 (10-36)</td>
<td>Beta</td>
<td>Based on the proportion of people likely to “take action” after a UK-based mass-media campaign to promote app use (Brannan et al., [13])</td>
</tr>
<tr>
<td>Users who adhered to physical activity app (weighted annual average), % (UI)</td>
<td>15 (10-21)</td>
<td>Beta</td>
<td>Weighted average of estimates of “app only” adherence from Guertler et al [23]</td>
</tr>
<tr>
<td>Intervention increase in physical activity for those who adhered to the app (mins/week), n (SD)</td>
<td>285 (43)</td>
<td>Normal</td>
<td>Reported increase of 1404 steps per day from recent meta-analysis of randomized controlled trials (Gal et al [8]) was converted to MVPA(^c)-MET(^d) mins/week, assuming a conversion factor of 34.5 steps equating to 1 MVPA-MET min [24]</td>
</tr>
<tr>
<td>Cost of a one-off national level mass media campaign (NZ $), n (SD %)</td>
<td>2,883,000 (20)</td>
<td>Gamma</td>
<td>As per a similar NZ study for promoting a weight loss app, by Cleghorn et al [19]; includes costs associated with identification of high-quality apps and mass media campaign across multiple media</td>
</tr>
</tbody>
</table>

\(^a\)NZ: New Zealand.  
\(^b\)UI: uncertainty interval.  
\(^c\)MVPA: moderate-to-vigorous physical activity.  
\(^d\)MET: metabolic equivalent of task.

The total cost of the intervention was estimated at NZ $2,883,000 (using consumer price index and purchasing power parity adjustments, the currency exchange rate used for this paper was 2011 NZ $1=2018 US $0.73) from a previous NZ study of the costs associated with a modeled mass media campaign to promote a weight loss app [19]. The cost of the intervention captured the costs associated with identifying the highest quality apps, promotion of the highest quality apps for...
physical activity on relevant government websites, and a mass media campaign rolled out across multiple media. These costs were also similar to the estimated cost for a modeled mass media campaign to promote smoking cessation apps [18].

Increases in physical activity were applied to the proportion of the population who downloaded and used the app for at least 7 days. We assumed that the intervention effect would apply to adults 15-79 years of age. This was the population range covered in studies included in the review used to estimate physical activity increases in response to physical activity apps [8].

For those who used the app for at least 7 days, we estimated that physical activity would increase by an average of 285 moderate-to-vigorous physical activity–metabolic equivalent of task (MET) mins/week using a recent meta-analysis examining increases in physical activity associated with app use [8]. This is equivalent to 1.6 hours of additional brisk walking per week. We assumed that the intervention increase in physical activity would wane over the course of the year in which the intervention was implemented, with no effect beyond the first year of the intervention. This was in line with the source of our estimate of intervention increase in physical activity, where included studies were evaluated based on the short-term (<3 months) impacts.

BAU was assumed to include the existing levels of physical activity and existing levels of physical activity app use, with no additional promotion. The current promotion of physical activity apps in New Zealand was considered negligible, and therefore, the BAU physical activity distribution was assumed to reflect the continuation of a low or no physical activity app promotion environment.

**Multistate Life Table Model**

The model consists of a main life table parameterized with age, sex, and ethnicity (Māori—the indigenous population of New Zealand—and non-Māori) specific all-cause mortality and morbidity rates. Alongside the main life table are 9 parallel physical activity and transport-related disease life tables where proportions of the population simultaneously reside: coronary heart disease (CHD), stroke, type 2 diabetes, colorectal cancer, breast cancer (females only), chronic obstructive pulmonary disease (COPD), lower respiratory tract infection (LRTI), lung cancer, and road transport injury. Modeled diseases include both physical activity and transport-related conditions, as the model was designed to examine both interventions. COPD, LRTI, lung cancer, and road transport injury were inactive (ie, “turned off”) in this study, as they are not associated with physical activity (see Figure 2 for the conceptual diagram, adapted from Mizdrak et al [20]). The proportions of the population in each disease life table at each annual time step are a function of past and current disease incidence, case fatality, and remission (for cancers only) rates.

**Figure 2.** Conceptual diagram of model. CHD: coronary heart disease; COPD: chronic obstructive pulmonary disease; LRTI: lower respiratory tract infection.

The physical activity distribution of the New Zealand adult population was estimated by converting responses to the New Zealand Physical Activity Questionnaire Short Form in the New Zealand Health Survey to MET minutes per week of moderate and vigorous physical activity. A MET is the ratio of work metabolic rate to a standard resting metabolic rate, where 1 MET is equivalent to quiet sitting [27]. Brisk walking was assigned a MET value of 3.0, moderate activities a MET value of 4.5, and vigorous activities a MET value of 6.5 [20].
The modeled intervention induced changes in physical activity were combined with relative risks for the association between physical activity and disease outcomes (CHD, stroke, type 2 diabetes, breast cancer, colorectal cancer) to produce population impact fractions [28]. These were used in the model to modify incidence rates of diseases, which in turn results in changes in all-cause mortality and morbidity rates. The model includes time lags to account for the nonimmediate impact of changes in population risk distribution on disease incidence: changes in CHD, stroke, and type 2 diabetes are based on the average population impact fraction over the past 0-5 years, for cancers on average for the previous 10-30 years [20]. For modeling parsimony, we assumed that there would be no impact of the modeled intervention on health beyond that captured through the diseases previously mentioned, including no impact on obesity, injury, or mental health outcomes. These assumptions are consistent with the evidence base: there does not appear to be a consistent association with weight loss for apps that specifically target physical activity [29], and there is no evidence (to our knowledge) of the impact of physical activity apps on mental health outcomes or injury (as covered further in the Discussion).

In addition, the model captures changes in health system costs associated with changing disease prevalence and population longevity. Disease-specific costs were based on the timing of events (first year, subsequent year, and last 6 months of life) and were derived according to an established protocol [30]. Changes in the proportion of the population in each disease state result in proportional changes in health system costs, and the model captures unrelated health system costs (ie, increases in health system costs out into the future due to people living longer as a result of the modeled intervention).

Our results project the health gains and health system cost impacts for the remainder of the life course of the modeled population. Both health gains and health system costs were discounted at 3%, with key results using 0% and 6% discount rates presented as sensitivity analyses. We also ran the results applying an “equity adjustment” that set background mortality and morbidity rates for Māori to non-Māori values, a routinely used modeling technique that avoids undervaluation of health gains for disadvantaged populations [31]. Scenario analyses included a scenario where the age range for the intervention was restricted to those 40-79 years of age with total intervention costs remaining the same, and one in which we assumed that the intervention impact would be maintained for 5 years following the intervention. Finally, tornado plots show the contribution of assumptions around each step in the intervention pathway to model uncertainty of the results.

The model was built in Microsoft Excel (Microsoft Corporation) and run using a macro written in Visual Basic for Applications. Uncertainty around health gains and cost-effectiveness was estimated using a Monte Carlo analysis; the model was run 2000 times with parameters sampled independently from their respective probability distributions. Results are given as the 50th percentile of all model runs, with 2.5th and 97.5th percentiles representing the 95% uncertainty interval (UI) around modeled values. The probability of cost-effectiveness at different monetary thresholds was based on the proportion of model runs with an incremental cost-effectiveness ratio (ICER) below the threshold. Further model details are provided in a technical report [20].

**Results**

The one-off mass media campaign promoting smartphone apps for physical activity resulted in an increase of 28 QALYs (95% UI 8-72) over the lifetime of the 2011 population, or 0.008 QALYs gained per 1000 people (see Table 2). The modeled improvements in health came at a net cost of NZ $2.2 million (US $1,625,000; 95% UI 1.02 million-3.5 million). The ICER was NZ $81,000 (US $59,000; 95% UI 17,000-345,000) per QALY gained. The intervention had a low probability (20%) of being cost-effective at a cost-effectiveness threshold of NZ $45,000 per QALY gained (see Figure 3).

Health gains per capita were higher in older age groups, and, assuming the intervention costs were spread evenly across the eligible population, the intervention was more likely to be cost-effective in older age groups compared to younger age groups (ie, more likely to be under the NZ $45,000 threshold). Health gains for Māori increased with the application of the “equity adjustment” (ie, non-Māori mortality and morbidity rates used for Māori; see Table 3).
Table 2. Health gains and cost-effectiveness of a mass media campaign to promote physical activity smartphone apps by age, sex, and ethnicity (lifetime gains, 3% discount rate).

<table>
<thead>
<tr>
<th>Sex, ethnicity</th>
<th>Age group (years)</th>
<th>QALYs(^a)/1000 population (UI(^b))</th>
<th>Cost per QALY gained: ICER(^c), 2011 NZ $ (UI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All, all</td>
<td>All groups</td>
<td>0.008 (0.002-0.021)</td>
<td>81,000 (17,000-345,000)</td>
</tr>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-Māori</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;40</td>
<td>0.001 (0.000-0.003)</td>
<td>606,000 (190,000-2,368,000)</td>
</tr>
<tr>
<td></td>
<td>40-60</td>
<td>0.008 (0.002-0.021)</td>
<td>86,000 (16,000-384,000)</td>
</tr>
<tr>
<td></td>
<td>60-80</td>
<td>0.021 (0.006-0.055)</td>
<td>27,000 (cost-saving(^d) to 147,000)</td>
</tr>
<tr>
<td></td>
<td>Māori</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;40</td>
<td>0.002 (0.001-0.006)</td>
<td>354,000 (111,000-1,384,000)</td>
</tr>
<tr>
<td></td>
<td>40-60</td>
<td>0.018 (0.005-0.047)</td>
<td>35,000 (2000-179,000)</td>
</tr>
<tr>
<td></td>
<td>60-80</td>
<td>0.031 (0.009-0.083)</td>
<td>16,000 (cost-saving to 96,000)</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-Māori</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;40</td>
<td>0.002 (0.000-0.005)</td>
<td>393,000 (120,000-1,499,000)</td>
</tr>
<tr>
<td></td>
<td>40-60</td>
<td>0.006 (0.002-0.017)</td>
<td>119,000 (26,000-495,000)</td>
</tr>
<tr>
<td></td>
<td>60-80</td>
<td>0.023 (0.007-0.061)</td>
<td>26,000 (cost-saving to 132,000)</td>
</tr>
<tr>
<td></td>
<td>Māori</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;40</td>
<td>0.003 (0.001-0.009)</td>
<td>196,000 (54,000-768,000)</td>
</tr>
<tr>
<td></td>
<td>40-60</td>
<td>0.019 (0.005-0.049)</td>
<td>31,000 (0-158,000)</td>
</tr>
<tr>
<td></td>
<td>60-80</td>
<td>0.035 (0.010-0.094)</td>
<td>15,000 (cost-saving to 87,000)</td>
</tr>
</tbody>
</table>

\(^a\)QALY: quality-adjusted life year.

\(^b\)UI: uncertainty interval.

\(^c\)ICER: incremental cost-effectiveness ratio.

\(^d\)Negative cost per QALY gained (ie, the intervention results in cost-savings to the health system).

**Figure 3.** Probability of the modeled physical activity app promotion intervention being cost-effective for different cost-effectiveness thresholds (in cost per quality-adjusted life year gained).
Table 3. Results for Māori (Indigenous population) with equity adjustment applied (lifetime gains, 3% discount rate).

<table>
<thead>
<tr>
<th>Sex, age (years)</th>
<th>QALYs/1000 people (UI)</th>
<th>Cost per QALY gained: ICER, 2011 NZ $ (UI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>0.002 (0.001-0.006)</td>
<td>315,000 (92,000-1,191,000)</td>
</tr>
<tr>
<td>40-60</td>
<td>0.022 (0.006-0.058)</td>
<td>30,000 (1000-142,000)</td>
</tr>
<tr>
<td>60-80</td>
<td>0.046 (0.012-0.126)</td>
<td>11,000 (cost-saving(^{d}) to 69,000)</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>0.004 (0.001-0.010)</td>
<td>172,000 (43,000-669,000)</td>
</tr>
<tr>
<td>40-60</td>
<td>0.024 (0.006-0.062)</td>
<td>26,000 (cost-saving to 130,000)</td>
</tr>
<tr>
<td>60-80</td>
<td>0.052 (0.014-0.137)</td>
<td>10,000 (cost-saving to 64,000)</td>
</tr>
</tbody>
</table>

\(^{a}\)QALY: quality-adjusted life year.
\(^{b}\)UI: uncertainty interval.
\(^{c}\)ICER: incremental cost-effectiveness ratio.
\(^{d}\)Negative cost per QALY gained (ie, the intervention results in cost-savings to the health system).

We explored the impact of selected changes to model specification on the results (see Table 4). Given that the intervention was least cost-effective in the youngest age group, we ran a scenario analysis to determine the extent that the overall cost-effectiveness might be improved by narrowing the population targeted by the intervention to those 40-80 years of age. This slightly increased the average cost-effectiveness of the intervention. Assuming the impact of the intervention held for 5 years rather than 1 year, the health gains would be over four times larger than in the main analysis and would result in much lower health system costs, resulting in a highly cost-effective ICER of NZ $2000 per QALY gained. Changing the discount rate had the expected impact on the overall results, with a zero-discount rate resulting in higher health gains. Finally, we examined the contribution of different intervention parameters to uncertainty in the modeled results. Uncertainty in health gains was driven by uncertainty in the app use parameter, and uncertainty in health system cost impacts was driven by uncertainty in the intervention cost parameter (see Figures S1 and S2 in Multimedia Appendix 1). The picture for the ICER was less clear; uncertainty in app use was the greatest contributor to uncertainty in the ICER, but this was closely followed by uncertainty around other intervention parameters (see Figure 4).

Table 4. Sensitivity and scenario analyses for a one-off national-level mass media campaign to promote smartphone apps for physical activity (expected value analysis, lifetime perspective, 3% discount rate, unless otherwise stated).

<table>
<thead>
<tr>
<th>Sensitivity/scenario analyses(^{a})</th>
<th>Health gain (QALY(^{c}))</th>
<th>Net health system costs (NZ $)</th>
<th>Cost per QALY gained: ICER(^{b}) (NZ $)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base case analysis</td>
<td>33</td>
<td>2,315,000</td>
<td>81,000</td>
</tr>
<tr>
<td>Target age range set to 40-80 years of age (otherwise base case)</td>
<td>30</td>
<td>2,387,000</td>
<td>80,000</td>
</tr>
<tr>
<td>5-year maintenance of additional physical activity levels followed by a return to preintervention levels (otherwise base case)</td>
<td>126</td>
<td>241,146</td>
<td>2000</td>
</tr>
<tr>
<td>0% discount rate</td>
<td>57</td>
<td>2,153,000</td>
<td>38,000</td>
</tr>
<tr>
<td>6% discount rate</td>
<td>22</td>
<td>2,332,000</td>
<td>108,000</td>
</tr>
</tbody>
</table>

\(^{a}\)Expected values given for all scenarios.
\(^{b}\)QALY: quality-adjusted life year.
\(^{c}\)ICER: incremental cost-effectiveness ratio.
Discussion

Principal Findings

We modeled the likely impact of a one-off national-level mass media campaign to promote uptake of smartphone apps for physical activity using published estimates of uptake, adherence, and effectiveness [8,13,23,25,26]. Modeled through changes in disease incidence the intervention has a 20% chance of being cost-effective for the whole target population at a commonly applied threshold of GDP per capita of the country (ie, NZ $45,000 per QALY gained for New Zealand) [32]. There was also wide uncertainty around the health system cost impacts and cost-effectiveness of the intervention.

Comparison With Prior Work

This is the first study of the cost-effectiveness of mass media promotion of smartphone apps for physical activity, at least that we are aware of. This study contributes to calls to build the evidence base on the cost-effectiveness of physical activity interventions [33]. This work also has a high level of comparability with previous research on other health-related app promotion in the NZ setting. We found that a mass media campaign to promote physical activity apps appears to be less effective in achieving health gain and less cost-effective than mass media campaigns to promote smoking cessation (modeled impact: 6760 QALYs, NZ $115 million savings to the health system [18]) but was similar to a campaign to promote the use of weight loss apps (modeled impact: 29 QALYs, ICER of NZ $79,700 [19]). This suggests that mass media campaigns to promote apps may have different impacts depending on the behavior targeted by apps.

Our results also indicate lower effectiveness and poorer cost-effectiveness on a per capita basis than previous research that modeled the effectiveness of a mass media campaign and other strategies to promote physical activity in Australia [34]. This is likely due to our study applying more conservative estimates for the impact of a mass media campaign intervention than earlier work and differences in underlying physical activity patterns and epidemiology across different countries.

Strengths and Limitations

This study has the strength of using an established multistate life table model based on rich disease-specific epidemiological and costing data. Multistate life table modeling captures health impacts across multiple diseases over time. The widespread use of this modeling methodology across Australasia means we are able to compare our results to those of other health interventions (eg, [18,19,34]). Limitations of multistate life table modeling include the assumption of disease independence and our use of a health-system perspective for costs and benefits. Regarding the former, we do account for the relationship between type 2 diabetes and CHD and stroke, given that type 2 diabetes is a risk factor for these conditions. The health system perspective of this study means that potential costs and benefits outside the health system (eg, the cost of the original development of physical activity apps) were not captured. However, our methods could be adapted to include additional costs and benefits for different audiences. For example, we are currently exploring the inclusion of productivity impacts, such as income loss from disease diagnosis, into our models.

Effect sizes were based on a review of relevant sources that were then assessed on both methodological rigor and appropriateness for modeling (see Multimedia Appendix 1 for further details). We also were able to model parameter uncertainty around all the key parameters and a range of sensitivity and scenario analyses.

We presented heterogeneity in the health impacts of the modeling intervention. As we assumed no heterogeneity in intervention impact, our estimates reflected differential health gains owing to underlying differences in physical activity levels and epidemiology, and not differential response to the intervention. There was insufficient information in the sources of parameter estimates to suggest differences in intervention impacts by age, sex, or ethnicity. Previous research has shown
high levels of engagement with physical activity apps across subpopulations, including older adults [35] and different ethnic groups [36]. However, if certain population groups are more or less likely to respond to the intervention, then this would influence the overall effectiveness and cost-effectiveness. Future evaluations of physical activity app interventions should explicitly consider differential impacts in intervention uptake and efficacy, especially given that our results showed differences in the health gain likely to be achieved.

We assumed that any impact of the intervention on physical activity levels would be restricted to the year in which the intervention was implemented, consistent with existing evidence. The majority of physical activity apps have only been evaluated for short-term (<3 months) impact [7,8]. Our estimate of average adherence in the year that the app was implemented is consistent with the small number of studies that have evaluated physical activity impacts beyond 3 months [23,37]. There was no evidence to support modeling an intervention impact that extended beyond 1 year, and this highlights the need for longer-term evaluations of physical activity app interventions, especially as our scenario analysis demonstrates that considerably larger health gains could be achieved if intervention impacts were maintained over time. Evidence of long-term impacts on physical activity have been observed with interventions to improve walking and cycling infrastructure [38,39], suggesting that structural interventions that change environments may be more effective than interventions targeting individual-level behavior change in an unsupportive environment.

The model captures health impacts through the conditions that are strongly associated with MVPA including cardiovascular diseases, type 2 diabetes, and selected cancers. We did not capture potential additional health gains or losses from obesity, mental health, or injury. Although apps have been shown to be effective in promoting weight loss [40], effect sizes predominantly capture apps designed to influence both physical activity and dietary behaviors. For apps that specifically target physical activity, there does not appear to be a consistent association with weight loss [29]. In addition, the impact of physical activity apps on mental health outcomes has not been quantified (to our knowledge). Although regular physical activity is associated with improved mental health [41], recent evidence suggests that a high percentage of fitness apps contain features linked to negative body image and maladaptive exercise behavior [42]. Further research is needed to understand whether health gains associated with increased physical activity through app use are complemented or counteracted by other health-related outcomes, including weight loss, injury, and mental health.

**Policy Implications**

The wide UIs around our modeled results demonstrate the need for better evaluations of app-based and other physical activity interventions. In particular, we need to better understand what interventions are most likely to result in long-term maintenance of physical activity increases, as these are the interventions that will result in the largest health gains. Modeling studies such as this one are a valuable approach to quantify the health gains that may be possible with different intervention options prior to implementing specific interventions.

Although physical activity apps offer the potential to increase physical activity at the individual level, our results suggest that promoting physical activity apps through mass media is currently unlikely to be an effective or cost-effective public health intervention at the population level, at least with existing app designs and mass media campaign methods. Cost-effectiveness of a mass media campaign to promote smartphone apps for physical activity could be improved by delivering more targeted campaigns using social media; this may deliver health gains at lower cost than the intervention modeled here. Worldwide, there is recognition that targeted, individual-focused interventions need to be combined strategically with policy actions that support physical activity [41]. Our results suggest that a mass media campaign to promote smartphone apps for physical activity is not cost-effective as a stand-alone intervention. Other strategies to promote physical activity that result in long-term behavior change are likely to be more effective (eg, investment in walking and cycling infrastructure [38]). Our results are likely to be generalizable to similar contexts—high-income countries with similar epidemiology, physical activity levels, app uptake, and other population characteristics.

**Conclusion**

A one-off national-level mass media campaign to promote the use of smartphone apps for physical activity is unlikely to generate much health gain. Based on current and often weak evidence, it also appears unlikely to be cost-effective at the population level. Investments in physical activity that are associated with long-term maintenance of behavior are likely to be of greater benefit.

**Acknowledgments**

The authors gratefully acknowledge the Health Research Council of New Zealand (HRC16/443) for providing the funding for this study.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Supplementary materials.
References


Abbreviations

BAU: business-as-usual
CHD: coronary heart disease
COPD: chronic obstructive pulmonary disease
ICER: incremental cost-effectiveness ratio
LRTI: lower respiratory tract infection
MET: metabolic equivalent of task
MVPA: moderate-to-vigorous physical activity
QALY: quality-adjusted life years
UI: uncertainty interval
Optimizing Smartphone-Delivered Cognitive Behavioral Therapy for Body Dysmorphic Disorder Using Passive Smartphone Data: Initial Insights From an Open Pilot Trial

Hilary Weingarden, PhD; Aleksandar Matic, PhD; Roger Garriga Calleja, MS; Jennifer L Greenberg, PsyD; Oliver Harrison, MA, MBBS, MPH; Sabine Wilhelm, PhD
1Massachusetts General Hospital/Harvard Medical School, Boston, MA, United States
2Telefónica Alpha, Barcelona, Spain

Corresponding Author:
Hilary Weingarden, PhD
Massachusetts General Hospital/Harvard Medical School
185 Cambridge Street
Suite 2000
Boston, MA, 02114
United States
Phone: 1 6176436206
Email: hilary_weingarden@mgh.harvard.edu

Abstract

Background: Smartphone-delivered cognitive behavioral therapy (CBT) is becoming more common, but research on the topic remains in its infancy. Little is known about how people typically engage with smartphone CBT or which engagement and mobility patterns may optimize treatment. Passive smartphone data offer a unique opportunity to gain insight into these knowledge gaps.

Objective: This study aimed to examine passive smartphone data across a pilot course of smartphone CBT for body dysmorphic disorder (BDD), a psychiatric illness characterized by a preoccupation with a perceived defect in physical appearance, to inform hypothesis generation and the design of subsequent, larger trials.

Methods: A total of 10 adults with primary diagnoses of BDD were recruited nationally and completed telehealth clinician assessments with a reliable evaluator. In a 12-week open pilot trial of smartphone CBT, we initially characterized natural patterns of engagement with the treatment and tested how engagement and mobility patterns across treatment corresponded with treatment response.

Results: Most participants interacted briefly and frequently with smartphone-delivered treatment. More frequent app usage ($r=-0.57$), as opposed to greater usage duration ($r=-0.084$), correlated strongly with response. GPS-detected time at home, a potential digital marker of avoidance, decreased across treatment and correlated moderately with BDD severity ($r=0.49$).

Conclusions: The sample was small in this pilot study; thus, results should be used to inform the hypotheses and design of subsequent trials. The results provide initial evidence that frequent (even if brief) practice of CBT skills may optimize response to smartphone CBT and that mobility patterns may serve as useful passive markers of symptom severity. This is one of the first studies to examine the value that passively collected sensor data may contribute to understanding and optimizing users’ response to smartphone CBT. With further validation, the results can inform how to enhance smartphone CBT design.

(JMIR Mhealth Uhealth 2020;8(6):e16350) doi:10.2196/16350

KEYWORDS
body dysmorphic disorder; cognitive behavioral therapy; mobile health; mobile phone; patient engagement

Introduction

Background

The supply and demand imbalance between those who need psychological treatment and those who are able to receive it represents a serious public health concern [1,2]. Indeed, only 43.6% of those with psychiatric illnesses in the United States receive treatment and fewer receive gold-standard treatment [2]. Moreover, certain psychiatric illnesses are less well-recognized than others, and under-recognized illnesses...
likely have the biggest access to care gaps. For example, 35.1% of adults with body dysmorphic disorder (BDD), a psychiatric illness characterized by a preoccupation with a perceived defect in physical appearance [3], receive psychotherapy; only 17.4% with BDD receive the gold-standard cognitive behavioral therapy (CBT) [4], despite strong research demonstrating its efficacy [5-7].

Fortunately, the development of smartphone-delivered CBT treatments may help address this access gap. Compared with in-person therapy, smartphone-delivered CBT is less expensive, more widely accessible, and highly flexible (eg, it can be used anywhere and anytime patients have their phones). The potential benefits of smartphone-delivered CBT are compounded by the growth of smartphone ownership. At present, 81% of the US population own a smartphone, a rate that has more than doubled since 2011 [8]. Not surprisingly, therefore, there is mounting enthusiasm among clinical researchers for developing and deploying smartphone CBT treatments [9,10].

Despite growing excitement, our understanding of smartphone-delivered CBT remains in its infancy, with a dramatic gap between the number of publicly available mental health apps and the paucity of scientific papers reporting on their evaluation [11]. In particular, very little is known about how people naturally engage with smartphone-delivered CBT compared with traditional in-person treatments, but it is likely that usage patterns differ dramatically. For example, in-person CBT is most commonly administered in once weekly, 50-min sessions, representing a concentrated and infrequent format. Next-generation internet-based CBT (ICBT) treatments, which have garnered substantial empirical support [12,13], are often built to mimic this style of longer duration, spaced out, formalized sessions because they were designed to be completed on one’s home computer. In both traditional CBT and ICBT, patients are instructed to practice skills between sessions to reinforce learning within real-world settings. Whereas the practice of skills between sessions has been associated with better CBT outcomes [14-16], many patients struggle to practice skills on their own between sessions. On the other hand, because people carry their phones at most times, smartphone-delivered treatments can be accessed by users at nearly any time and place. Having smartphone-delivered support available at all times may encourage practicing CBT skills with greater frequency and in a wider variety of settings than traditional in-person CBT, potentially opening doors to highly distinct engagement patterns. However, to date, we know very little about how often, for how long, or where people naturally engage with smartphone-delivered CBT treatments.

Moreover, very little is known about which engagement patterns correspond with an optimal response to smartphone-delivered CBT. Understanding optimal engagement patterns can allow for the design of more potent treatments by seeking to promote the most effective patterns of CBT app use. For example, gaining information about whether one’s frequency of use or duration of use matters more in terms of treatment response can inform whether apps should be designed to promote bursts of brief engagement or longer, less frequent sessions.

Finally, little is currently known about how the mobility patterns of patients change over the course of smartphone-delivered CBT. Previous research suggests that time spent at home, measured via a GPS, can serve as a digital marker of avoidance [17] and may correlate with symptom severity in depressive disorders [18]. Therefore, obtaining initial information about how mobility patterns change across smartphone treatment, and how these changes correspond with changes in severity, can inform treatment optimization by passively detecting changes in severity and triggering just-in-time interventions.

Altogether, in the field’s current, early stage of developing smartphone-delivered CBT treatments, we can benefit from examining pilot engagement and mobility data, to shape how we design optimal digital services and their clinical trials in the future. Smartphones offer a unique avenue for gaining rich insights into patterns of treatment engagement and predictors of treatment response because smartphones can unobtrusively (ie, in the background, without user input) collect a wide variety of sensor-based data over the course of treatment. For example, with patient consent, smartphones can be configured to passively collect objective information about patients’ engagement with the app (ie, how often and for how long patients use the program) as well as patients’ behavioral patterns over the course of treatment (eg, where patients typically use the app, changes in mobility patterns across treatment, via GPS). Passive data offer notable strengths for learning how to optimize smartphone-delivered treatments compared with more traditional assessment methods such as clinician interviews and self-reports. Passive smartphone data are sampled at a far greater frequency than traditional clinical assessments, which, at most, might be administered weekly. Frequent assessment that is conducted as one lives daily life captures richer contextual information, has higher temporal resolution to detect changes in symptoms or severity, and reduces the influence of recall biases that arise from subjective recollection of experiences over a broad time frame [19]. Altogether, passive smartphone data can offer valuable, low-burden insights into patterns of treatment engagement and digital markers of progress or deterioration, to optimize future design and research of smartphone-delivered treatments [20].

Objectives
To this end, this study exploratorily examines passive smartphone data from a 12-week open pilot trial of a smartphone-delivered CBT (Perspectives) for patients with BDD (N=10) to inform the study design, variables of interest, and hypothesis generation for future trials of smartphone-delivered CBT services. First, we aimed to initially characterize typical patterns of engagement with smartphone-delivered CBT for BDD in our sample, to obtain a preliminary understanding of how engagement may be similar to or different from participation in traditional in-person CBT. Second, we aimed to initially test how patterns of engagement corresponded with treatment response to inform early hypotheses about how we may design apps to optimize engagement and response. Third, we aimed to initially characterize the mobility patterns of participants across treatment, to preliminarily test whether GPS-based mobility patterns could serve as a digital marker of disorder severity. If validated in larger trials, digital...
markers of severity could be used to enhance treatments by triggering just-in-time interventions.

**Methods**

**Participants and Recruitment**

A paper by Wilhelm et al [21] gives detailed information on study methods, including a Consolidated Standards of Reporting Trials diagram, participant demographic information, and a description of the smartphone-delivered CBT for BDD treatment (ClinicalTrials.gov Identifier: NCT03221738).

A total of 10 adults with a primary psychiatric diagnosis of BDD were enrolled nationally in the open pilot trial (female: n=8, male: n=2; mean age 27.6, SD 5.66 years). Other inclusion criteria required that participants had at least moderately severe BDD symptoms (defined as a Yale-Brown obsessive compulsive scale modified for BDD [BDD-YBOCS] score >20), an acuity level appropriate for an outpatient level of care and lived in the United States. Exclusion criteria prohibited participation if the individual had a current severe major depressive disorder; borderline personality disorder; substance use disorder or acute, active suicidal ideation; had a lifetime diagnosis of bipolar disorder or a psychotic disorder; had cognitive impairment or intellectual disability that would interfere with participation; had engaged in previous CBT for BDD, or did not own an iPhone that supported the app software. Participants were either unmedicated or those on medication were required to be on a stable dose for at least two months before starting the study and were instructed not to change their medication regimen during the trial.

**Procedures**

Procedures were approved by the hospital’s institutional review board, and participants provided informed consent before beginning the study. Informed consent included a description of each type of passive smartphone data to be collected, a description of how those data were securely transmitted and deidentified before storage, the rationale for collecting those data, and a description of who would have access to the data.

**Assessments**

Clinical assessments were conducted by reliable, independent evaluators with a Master’s degree or doctorate, who were trained in primary diagnostic and outcome measures. Assessments for this study were conducted at the screening and baseline (same visit; week 0), midpoint (week 6), and posttreatment (week 12) assessments, and participants were compensated US $25 for completing the week 6 and week 12 assessments. Clinician-administered measures were collected via secure video calls that were Health Insurance Portability and Accountability Act (HIPAA) compliant. Self-report data were collected via Research Electronic Data Capture [22], a secure, HIPAA-compliant web-based survey collection platform.

In addition to providing clinical and outcome data, participants also provided qualitative feedback on the CBT app at several time points across the study. Specifically, written feedback was collected at the posttreatment assessment; oral feedback was gathered by members of the design team via separate interviews conducted shortly after the baseline, midpoint, and posttreatment clinical assessments.

**Treatment**

Following the screening and baseline assessment, the study staff instructed eligible participants on how to download and activate the Perspectives app onto their personal smartphones. The 12-week treatment consisted of psychoeducation and self-paced interactive exercises presented in a fixed order, which taught each of the core CBT skills for BDD (ie, cognitive restructuring, exposure with ritual prevention, mindfulness and perceptual retraining, core beliefs and self-esteem, engagement in value-based activities, and relapse prevention). The treatment was delivered via the smartphone app and was supported by light-touch communication with a doctoral-level therapist, whose primary role was to enhance motivation, address roadblocks, and answer questions [21]. Note that in this trial, Perspectives was developed for iPhones only; in 2018, iPhone operating systems represented approximately 44% of smartphones in the United States [23].

**Passive Smartphone Data Collection**

**Perspectives** was designed to passively collect information about app usage and mobility patterns of participants via GPS (the Measures section gives further details). We chose to collect these 2 types of passive data based on previous literature that points to their utility. In particular, app usage data may offer valuable insights into which engagement patterns are optimal for promoting treatment response [20], whereas mobility patterns from GPS can detect the proportion of time spent at home, a potential digital marker of avoidance [17]. As BDD is characterized by substantial avoidance (including housebound avoidance) [24], mobility patterns, therefore, have the potential to passively detect signs of symptom severity. By carefully selecting data categories and sampling rates (by default, the location was sampled whenever location changed by at least 100 m), the app was optimized to balance battery life and allowance of natural phone use. To this end, no participants complained about battery problems during the study.

**Measures**

**Clinical Assessments**

The Mini-International Neuropsychiatric Interview (version 7.0.2) [25] is a semistructured, clinician-administered diagnostic assessment of psychiatric illnesses. It was administered at the screening assessment to evaluate the inclusion and exclusion criteria.

The BDD-YBOCS [26] is a semistructured, clinician-administered, gold-standard assessment of current BDD symptom severity. The BDD-YBOCS is a 12-item Likert scale. Total scores range from 0 to 48, with higher scores corresponding to greater BDD severity. The BDD-YBOCS has strong psychometric properties, including internal consistency, interrater reliability, and test-retest reliability [26,27]. The BDD-YBOCS was administered at each assessment to evaluate the eligibility criteria (at screening) and changes in BDD severity. Percentage improvement in severity, a primary outcome in this study, was computed by dividing the difference between
baseline and posttreatment (week 12) BDD-YBOCS scores by the baseline value.

**Passive Smartphone Features**

To quantify and analyze the patterns of engagement with Perspectives and mobility across treatment, we computed several variables based on passive smartphone data.

**Quantity of App Use**

The quantity of app use was calculated as the total duration in minutes that a participant used the app. This was calculated by adding together all app sessions, or the periods of on-app time devoted to the therapy. Before analyses, together with designers of the Perspectives app, we considered various cutoff points for outliers in session length. Taking into account the possibility that participants might occasionally engage in multiple longer components of the app in sequence (eg, a mindfulness audio exercise, responding to coach messages, and completing an exposure exercise), we decided *a priori* on a session length cutoff of approximately 60 min, and outliers beyond this length were removed. To account for bursty use (ie, multiple brief usages separated by short breaks of <60 min in between), app usages that were separated by <60 min were summed together into a single session. For example, a participant who used the app for two 10-min increments with a 5-min break in between would be logged as having one 20-min session during this span. Quantity of app use was computed for the first half (6 weeks) and for the full 12 weeks of the CBT program (Table 1).

**Frequency of App Use**

This metric measured the extent to which a participant tended to use the app frequently or infrequently, expressed as the mean duration between 2 consecutive sessions, or periods of uninterrupted use. Frequency of app use was computed for the first half (6 weeks) and the full 12 weeks of the CBT program (Table 1).

**Mobility Patterns**

Using GPS data, we calculated the percentage of time spent at home during 1-week time intervals that overlapped with baseline, midpoint, and posttreatment BDD-YBOCS assessments (including 3 days before, 3 days after, and the day of BDD-YBOCS administration). Of note, at baseline, the BDD-YBOCS was typically administered on the same day the app was installed. Therefore, GPS data were not generally available for the 3 days before the baseline BDD-YBOCS assessment. Home location was inferred as the most common location ID captured between 3 AM and 6 AM per individual. All the remaining location IDs were labeled as *outside of home*. The various locations of participants were collected in a privacy-preserving way; each location where a participant spent at least 30 min was assigned a unique and random location ID (eg, ID78) and stored in the logs. This procedure was performed locally on the phone, and raw locations were removed before transferring the data to the server. The GPS sampling rate was set to 15 min, yet GPS readings were missing for 60% of the days.

**Statistical Analyses**

Data were analyzed using Python 3.6 (Python Software Foundation).

**Descriptive Patterns of App Usage**

To characterize the overall patterns of app usage, we visually inspected longitudinal patterns of usage by the participants across the 12-week treatment and we calculated the number of times the participants were engaged with the app for different lengths of time (ie, session durations). We elected not to identify subsamples based on usage (ie, clusters of users with similar engagement patterns) either visually or quantitatively, because of the small sample size.

**App Usage Patterns as Correlates of Percentage Improvement in the Yale-Brown Obsessive Compulsive Scale Modified for Body Dysmorphic Disorder**

To examine how the patterns of engagement of participants with Perspectives corresponded with their percentage improvement in BDD severity, we focused on 2 types of app usage patterns: quantity of app usage and frequency of app usage across the treatment. Normality was tested using the Shapiro-Wilk test and visual inspection. As the frequency of app use variable followed a long-tail distribution, log-transformation was performed before the analysis.

Two bivariate correlations were conducted, to preliminarily explore the relationships between the variables measuring (a) quantity and (b) frequency of app usage with percentage improvement in BDD-YBOCS from the baseline to week 12. Next, to initially examine the relative effect of quantity versus frequency of app use, a regression analysis of percent improvement was conducted, with both quantity and frequency of app use as independent variables. We primarily evaluated effect sizes, as opposed to statistical significance, for correlation and regression analyses, given the pilot nature of the data.

**GPS Data as a Correlate of the Yale-Brown Obsessive Compulsive Scale Modified for Body Dysmorphic Disorder Scores**

The relationship between symptom severity and mobility was explored via a bivariate correlation between BDD-YBOCS scores and the percentage of time spent at home during the week the BDD-YBOCS was measured. Note that absolute BDD-YBOCS scores were used for this analysis instead of percentage improvement, given the goal of exploring the predictive power of a GPS marker in assessing the current acuity of participants. The correlation analysis included 30 pairs of location variables and BDD-YBOCS scores (ie, 3 per participant, at baseline, midpoint, and posttreatment); thus, each participant was equally represented in the correlation analysis. Given that this analysis included multiple time points per participant, we followed up with a secondary analysis to verify that the results were not inflated based on the longitudinal nature of the data. Namely, 6000 correlation analyses were run by randomly selecting 1 of the 3 time points per participant (pre-, mid-, or posttreatment). This approach resulted in a very similar median correlation value to the analysis with 3 time points per participant; thus, secondary results are not presented. Again,
we primarily evaluated the effect size, as opposed to statistical significance, for this correlation analysis, to best account for the pilot nature of the study.

**Results**

Wilhelm et al [21] report the feasibility and acceptability of Perspectives, as well as the symptom improvement from baseline to posttreatment.

**Descriptive Patterns of App Usage**

We visually examined the longitudinal patterns of engagement with Perspectives across the 12-week treatment (Figure 1).

Despite the diversity in app usage across participants, several common usage patterns also emerged. First, most participants used the app with higher and lower intensities in the first and last weeks of the treatment, respectively (Figure 1). Additionally, data from both the app usage logs and GPS revealed that—within participants—participants generally preferred using the app at home over the first 8 weeks (1040/1488, 69.90% at home on average). During the ninth and tenth weeks, the proportion of app use at home and outside of home became more evenly distributed (60/105, 56.9% at home), and in the final 2 weeks of treatment, participants predominantly used the app outside of home (with only 16/98, 17% at home).

Moreover, unlike in-person therapy, most interactions with Perspectives were frequent (Figure 1) and very brief (Multimedia Appendix 1). The majority (374/510, 73.3%) of app sessions lasted ≤5 min. In only 11.7% (60/510) of cases, the app was used in sessions lasting 5 to 10 min, followed by 7.1% (36/510) and 6.4% (33/510) of cases in which the content was accessed for 10 to 20 or 20 to 40 min, respectively. Longer app usage was registered in only 1.4% (7/510) of the sessions. This pattern of brief engagement is consistent with how participants described their app usage in qualitative feedback. For instance, participants reported that they used the app during “dead time” while waiting (eg, in line at the store) or “for a few minutes each day to keep the lessons in mind” and described the app as “fast and easy to fit into your busy schedule.”

**App Usage Patterns as Correlates of Percentage Improvement in the Yale-Brown Obsessive Compulsive Scale Modified for Body Dysmorphic Disorder**

Means, standard deviations, and bivariate correlations of the percentage improvement in the BDD-YBOCS, the quantity of app usage, and frequency of app usage are provided in Table 1.
Table 1. Descriptive statistics and correlations between patterns of engagement with smartphone-delivered cognitive behavioral therapy for body dysmorphic disorder and treatment response.

<table>
<thead>
<tr>
<th>App usage patterns</th>
<th>Midtreatment, mean (SD)</th>
<th>Posttreatment, mean (SD)</th>
<th>Correlation with percentage improvement in BDD-YBOCS(^a) (baseline to week 12)(^b)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage improvement in BDD-YBOCS (%)</td>
<td>32.3 (23.8)</td>
<td>45.3 (14.7)</td>
<td>N/A(^c)</td>
<td>N/A</td>
</tr>
<tr>
<td>Quantity of app use</td>
<td>293.9 (276.4)</td>
<td>398.0 (310.25)</td>
<td>-0.084</td>
<td>.82</td>
</tr>
<tr>
<td>Frequency of app use(^d)</td>
<td>6.24 (0.7)</td>
<td>6.44 (0.73)</td>
<td>-0.57</td>
<td>.08</td>
</tr>
</tbody>
</table>

\(^a\)BDD-YBOCS: Yale-Brown obsessive compulsive scale modified for body dysmorphic disorder.
\(^b\)Correlations calculated for quantity and frequency of app use were derived from full 12-week treatment.
\(^c\)N/A: not applicable.
\(^d\)Frequency of app use values are log-transformed. Nontransformed means at midtreatment and posttreatment are 512 min (SD 1 255 to 1032 min) and 626 min (SD 1 301 to 1299 min), respectively.

The quantity of app usage was uncorrelated with percentage improvement in the BDD-YBOCS, whereas the frequency of app usage correlated strongly with treatment response and trended toward significance. The strong, negative relationship between mean (log) length of breaks between sessions (ie, frequency of app use) and improvement in the BDD-YBOCS initially suggests that shorter breaks between sessions corresponded with greater improvements (Table 1).

To follow up on patterns elucidated in bivariate correlations, we used regression analysis to preliminarily examine whether the frequency of app usage corresponded with treatment response more so than the quantity of app usage. When the primary outcome (percentage improvement in the BDD-YBOCS) was entered as a dependent variable, the frequency of app usage (ie, mean (log) duration between 2 consecutive sessions) predicted percentage improvement in the BDD-YBOCS with a small effect (\(\beta=-0.13; P=.03; 95\% \text{ CI} -0.231 \text{ to } -0.019\)), whereas the total quantity of app usage during the 12-week treatment did not predict improvement in the BDD-YBOCS (\(\beta=-0.08; P=.13; 95\% \text{ CI} -0.184 \text{ to } 0.027\)).

GPS Data as a Correlate of the Yale-Brown Obsessive Compulsive Scale Modified for Body Dysmorphic Disorder Scores

We used a scatterplot to visually inspect the relationship between time spent at home (based on GPS data) and symptom severity (measured with the BDD-YBOCS; Figure 2). The plot indicates that a shift occurred from baseline to posttreatment, characterized by a corresponding decrease in time spent at home and symptom severity. A follow-up correlational analysis suggests a moderately strong association between time spent at home and BDD symptom severity (\(r=0.49; P=.005\)).

Figure 2. Body dysmorphic disorder severity and time spent at home across treatment. BDD-YBOCS: Yale-Brown obsessive compulsive scale modified for body dysmorphic disorder.
Discussion

Principal Findings

Although enthusiasm for smartphone-delivered CBT is growing rapidly, there has not yet been substantial research on ways to enhance smartphone treatment. Before the widespread development and deployment of smartphone CBT treatments, it is important to first examine pilot data that characterizes the natural engagement patterns of users with smartphone-delivered CBT and identifies which usage and mobility patterns may optimize treatment. Such pilot data will provide timely information to researchers about variables and hypotheses of focus, in advance of larger, more costly validation trials, and can elucidate how we may explore enhancing smartphone-delivered CBT for optimum response in larger trials.

In particular, passively collected usage and sensor data from smartphones offer a unique, low-burden approach for gaining these important insights. Although a variety of passive data (eg, typing speed, activity level, phone usage, acoustic level) can be collected by smartphones [28], collecting sensor data involves a trade-off between gaining potentially useful information and depleting phone battery life (as well as risking user trust when collecting unnecessary data). Thus, initial signals from pilot research can shed light on which variables may be more or less fruitful to collect in clinical trials. To this end, this study used passive data from an open pilot trial of smartphone-delivered CBT for BDD, with the aim of preliminarily (1) characterizing the patterns of app usage of participants, (2) examining usage patterns that correspond with treatment response, and (3) examining mobility patterns that correspond with symptom severity.

Although app usage patterns varied substantially across participants, visual examination and descriptive analysis of usage data revealed several common patterns of engagement in our sample. First, participants tended to use the app more frequently and for a greater overall duration at the beginning of the 12-week treatment, with considerably lower usage later in treatment. This result is not surprising and may reflect that early on, participants required more time on the app to learn new information and skills. Later in the treatment, the participants may have transitioned to practicing greater applied skills, offline and in the real world [29]. In fact, qualitative feedback reflects that once participants learn skills, they practice them offline. For example, one participant reported, “I use the exercises all the time without the app. I have the big picture view of what I am trying to do.” Learning to use the treatment skills offline is likely an effective way to engage with smartphone-delivered CBT over time, as ultimately (like with in-person CBT), we hope for patients to internalize skills well enough to use them naturally as symptoms arise. Similarly, the results could reflect that participants simply received the necessary dose of treatment in a shorter time than the allotted 12 weeks [29]. On the other hand, lower usage at the end of treatment may reflect drops in engagement unrelated to CBT mastery (eg, because of boredom, lack of new content, loss of motivation). One participant’s posttreatment qualitative feedback supports this hypothesis; the participant reported that toward the end of the 12 weeks, there was less new material, and the participant was therefore not on the app as often. Reduced engagement over time is a very common challenge for app-based treatments [30]. Additional research is needed to fully understand the reasons for the reductions in app usage over time.

Second, descriptive results highlighted that participants typically used the app at home during the first two-thirds of treatment; later, the participants tended to use the app more when out of the house. This within-person pattern of increased usage outside of the house over time is consistent with the hypothesis that as participants gained CBT skills across treatment, they may have transitioned to using those skills offline and in the real world.

Finally, we observed that overall, the participants tended to use the app in brief and frequent sessions. In fact, most app sessions lasted <5 min each. This pattern reflects the way in which most people use smartphones in general: engaging with them often during short moments of downtime throughout the day [31]. This pattern also aligns with how we designed the app to be used. That is, we intentionally pared down content into brief text and exercises that could be completed quickly and repeated as often as one wished.

On the other hand, this pattern of brief and frequent sessions is notably distinct from how patients engage with face-to-face CBT or ICBT. Given the distinctive pattern of engagement we observed compared with better-established CBT modalities, it is critical to examine whether the naturally brief usage patterns of participants with smartphone-delivered CBT are effective or whether longer sessions are needed for response. Interestingly, preliminary correlation and regression results suggest that more frequent app usage, as opposed to greater duration of app usage, correlated strongly with treatment response—and trended toward statistical significance—in our (albeit small) sample. Consistent with these results, a previous review showed that overall time spent on web-based treatments for depression does not typically correlate with response to treatment [32]. In line with the aforementioned hypothesis that participants often practiced skills offline once learned, it is possible that the total duration of app usage does not fully capture the time participants spent engaging in treatment skills. Altogether, the results provide early, novel evidence that frequent (even if brief) practice of CBT skills may optimize the smartphone-delivered CBT response.

It is possible that frequent doses of practice help with learning CBT skills, as regular reinforcement of skills across broad contexts may enhance consolidation and generalization [33]. Researchers who are in the process of designing clinical trials to test smartphone-delivered CBT should consider collecting both quantity and frequency usage metrics to further validate optimum usage patterns. If validated in subsequent trials, the results have implications for the design of smartphone-delivered CBT. For example, findings suggest that information should be provided in brief chunks, as opposed to packing long, self-help—style psychoeducation into smartphone-delivered treatments. Moreover, it may be beneficial to design apps that are discreet, to promote frequent app use not only at home but also as symptoms arise in day-to-day life. App design can...
actively promote frequent use by incorporating reminders or rewards for use, in addition to including instructions to engage with the app often. Future research could test these design strategies using experimental designs to investigate which are effective for promoting frequent use.

In addition to usage patterns, we also examined mobility patterns from GPS data that correspond to BDD severity. Preliminary results showed that across treatment, the proportion of time spent at home—a potential digital marker of avoidance [17]—decreased. Time spent at home correlated positively with BDD severity across treatment, with a medium-to-large effect. Whether the proportion of time at home is truly tapping into avoidance behaviors (versus other aspects of BDD severity) is speculative and requires validation through future research. This is the first study to examine the time spent at home in relation to BDD severity. Whereas previous research has documented a link between time spent at home and depressive symptoms [18], because of the small sample, we did not examine this relationship when controlling for depression severity. However, as depression severity did not decrease across treatment in this sample [21], it is unlikely that the observed link is better accounted for by changes in depressive symptoms. Future research in a larger trial could parse apart the degree to which time spent at home serves as a digital marker of depressive versus BDD severity.

Altogether, strong initial GPS results underscore one variable where gains of data collection may outweigh costs; researchers designing upcoming smartphone-delivered CBT trials should consider measuring time spent at home, to further validate this potential unobtrusive marker of clinical severity. With further validation, detecting changes in one’s time spent at home could enhance smartphone-delivered CBT by unobtrusively triggering just-in-time interventions—a promising yet underdeveloped area of research [34]. For example, upon detecting increases in time spent at home, smartphone-delivered CBT treatments could send notifications to the user that reflect this observation (eg, “It looks like you’ve been spending more time at home”) and suggest adaptive strategies (eg, “Would you like to schedule an activity with a friend?”). Moreover, in larger trials, researchers can explore the utility of applying machine learning methods to predict changes in BDD severity from GPS-derived time spent at home.

Limitations

Results from this study should be interpreted, bearing in mind its limitations. Most notably, this pilot study had a small sample size. Thus, it is possible for 1 or 2 participants’ outlying usage patterns to unduly influence the results. That said, Kazdin [35] outlines a strong rationale for the ability to meaningfully examine data from small samples when data are collected at multiple time points across the treatment. Given the small sample size, we limited the scope of our aims and analyses to an exploratory examination of select patterns of interest, and we focused on robust effects that may indicate meaningful signals to follow-up. Follow-up in a larger sample would provide an opportunity to reliably test for statistical significance. To this end, results are intended to hone researchers’ decisions (eg, variables and hypotheses of focus) in advance of larger, more costly clinical trials of smartphone-delivered CBT treatments rather than to provide conclusive evidence in and of themselves.

In addition to a small sample, this pilot trial specifically focused on smartphone-delivered CBT for BDD. It is possible that insights will not generalize to smartphone-delivered CBT treatments for other disorders. However, given the core similarities between CBT for BDD and many other psychiatric conditions, such as anxiety disorders, obsessive-compulsive–related disorders, and eating disorders, we anticipate that findings will be relevant in the design of smartphone-delivered CBT treatments for related conditions. Finally, our strong initial GPS results should be interpreted, bearing in mind the high degree of missing GPS data (683/1134, 60.23% of the days) in our sample. Although the specific reasons for missing GPS data in our study are unknown, a high rate of missing geolocation data in mobile research is typical (eg, ranging from 40% to 90% missing) [36-39] and may be attributed to a range of factors, including participants switching off the device, participants activating a mode that does not permit location services (eg, airplane mode) or problems with permission to access the location sensor that can occur with the iPhone platform [36]. Importantly, missing GPS data in our study did not correlate with the BDD symptom severity of the participants and therefore were likely random with respect to BDD symptoms. Thus, it is unlikely that patterns of missingness meaningfully influenced this correlation result. As with other results in this pilot study, these initial findings should be used for hypothesis generation at this stage.

Conclusions

This study also had several notable strengths. First, whereas many existing smartphone-delivered CBT trials use nonclinical or convenience samples, we used a clinical sample that was diagnosed and assessed via gold-standard, clinician-administered measures. Participants were recruited nationally, which may enhance the generalizability of our initial findings. Finally, the correlation results for app usage and GPS patterns were robust despite our small sample, suggesting that these novel insights have strong potential to enhance costly, well-powered future trials.

Altogether, the results suggest that as researchers design efficacy trials to test smartphone-delivered CBT, it is worthwhile to collect data on patterns of use (with a focus on frequency versus quantity of use) and time spent at home. Novel study results suggest that these variables may correspond meaningfully with the response to treatment and, with further validation, may inform how to enhance smartphone-delivered CBT interventions.
Acknowledgments
The authors would like to thank Remko Vermeulen, Anna Hiltunen, Thomas McCoy, Berta Summers, and Ilana Ladis for their work on this study. SW served as the principal investigator and designed the trial, with HW and JG. HW and AM conceptualized this paper. HW and AM wrote the manuscript. AM and RC conducted the statistical analyses. All authors read, edited, and approved the manuscript before submission. This work was supported by Telefonica Innovation Alpha.

Conflicts of Interest
The sponsor and investigators from Massachusetts General Hospital collaborated in the development of the Perspectives digital service, and they collaborated on the analysis and writing of this manuscript. HW, JG, and SW received salary support from Telefonica Innovation Alpha and are presenters for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies. SW has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, and Oxford University Press. SW has also received speaking honoraria from various academic institutions and foundations, including the International Obsessive Compulsive Disorder Foundation and the Tourette Association of America. In addition, SW received payment from the Association for Behavioral and Cognitive Therapies for her role as Associate Editor for the Behavior Therapy journal as well as from John Wiley & Sons, Inc for her role as Associate Editor on the journal Depression & Anxiety. OH, RG, and AM are employees of Telefonica Innovation Alpha.

Multimedia Appendix 1
Distribution of session durations aggregated across the sample.

References


Abbreviations

- BDD: body dysmorphic disorder
- BDD-YBOCS: Yale-Brown obsessive compulsive scale modified for body dysmorphic disorder
- CBT: cognitive behavioral therapy
- HIPAA: Health Insurance Portability and Accountability Act
- ICBT: internet-based cognitive behavioral therapy

©Hilary Weingarden, Aleksandar Matic, Roger Garriga Calleja, Jennifer L Greenberg, Oliver Harrison, Sabine Wilhelm. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 18.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Evaluation of the Tobbstop Mobile App for Smoking Cessation: Cluster Randomized Controlled Clinical Trial

Merixtell Pallejà-Millà1,2*, MSD; Cristina Rey-Reñones2,3*, NP, PhD; Maria Luisa Barrera Uriarte3,4, NP; Esther Granado-Font3,5, NP, PhD; Josep Basora1,2, MD; Gemma Flores-Mateo1,6, MD; Jordi Duch1, PhD

1Unitat de Suport a la Recerca Camp de Tarragona, Fundació Institut Universitari per a la recerca a l’Atenció Primària de Salut Jordi Gol i Gurina, Reus, Spain
2Departament de Ciències Mèdiques Bàsiques, Universitat Rovira i Virgili, Reus, Spain
3Institut Català de la Salut, Unitat de Suport a la Recerca Camp de Tarragona, Reus, Spain
4Equip d’Atenció Primària La Granja (Tarragona-2), Direcció d’Atenció Primària Camp de Tarragona, Institut Català de la Salut, Torreforta, Spain
5Equip d’Atenció Primària Horts de Miró (Reus-4), Direcció d’Atenció Primària Camp de Tarragona, Institut Català de la Salut, Reus, Spain
6Unitat d’Anàlisi i Qualitat, Xarxa Sanitària i Social Santa Tecla, Tarragona, Spain
7Departament d’Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain
‡Grup de Recerca Tecnologies de la Informació i Comunicació en Atenció Primària
*these authors contributed equally

Corresponding Author:
Cristina Rey-Reñones, NP, PhD
Institut Català de la Salut
Unitat de Suport a la Recerca Camp de Tarragona
C/ Camí de Riudoms, 53
Reus, 43202
Spain
Phone: 34 977778515 ext 405
Email: crey.tgn.ics@gencat.cat

Abstract

Background: Mobile apps provide an accessible way to test new health-related methodologies. Tobacco is still the primary preventable cause of death in industrialized countries, constituting an important public health issue. New technologies provide novel opportunities that are effective in the cessation of smoking tobacco.

Objective: This paper aims to evaluate the efficacy and usage of a mobile app for assisting adult smokers to quit smoking.

Methods: We conducted a cluster randomized clinical trial. We included smokers older than 18 years who were motivated to stop smoking and used a mobile phone compatible with our mobile app. We carried out follow-up visits at 15, 30, and 45 days, and at 2, 3, 6, and 12 months. Participants of the intervention group had access to the Tobbstop mobile app designed by the research team. The primary outcomes were continuous smoking abstinence at 3 and 12 months.

Results: A total of 773 participants were included in the trial, of which 602 (77.9%) began the study on their D-Day. Of participants in the intervention group, 34.15% (97/284) did not use the app. The continuous abstention level was significantly larger in the intervention group participants who used the app than in those who did not use the app at both 3 months (72/187, 38.5% vs 13/97, 13.4%; P<.001) and 12 months (39/187, 20.9% vs 8/97, 8.25%; P=.01). Participants in the intervention group who used the app regularly and correctly had a higher probability of not being smokers at 12 months (OR 7.20, 95% CI 2.14-24.20; P=.001) than the participants of the CG.

Conclusions: Regular use of an app for smoking cessation is effective in comparison with standard clinical practice.

Trial Registration: Clinicaltrials.gov NCT01734421; https://clinicaltrials.gov/ct2/show/NCT01734421

(JMIR Mhealth Uhealth 2020;8(6):e15951) doi:10.2196/15951

KEYWORDS

tobacco smoking; tobacco use cessation; mobile application; primary public health; clinical trial; mobile phone
Introduction

The number of health interventions involving mobile-based technologies has been increasing in recent years, given the potential of reaching larger audiences who own and use smartphones on a daily basis. In the case of Spain, between 80% and 90% of the population have at least one smartphone [1], which places the country at the top of European mobile phone usage, with 23 million people owning smartphones [2]. Mobile apps provide an accessible way to test new health-related methodologies, which also address user concerns around the availability and confidentiality of their personal data [3]. In Spain, users download around 4 million apps every day and, more importantly, two-thirds of teenagers (and young users) downloaded and used a mobile health app in the last year.

Tobacco consumption levels around the world vary significantly; while in some low-income countries there has been an increase in the prevalence of tobacco, in industrialized countries the observed decrease in consumption in recent years seems to have stopped. Nonetheless, tobacco is still the primary preventable cause of death in industrialized countries, constituting an important public health issue despite all the medical advances and resources invested to reduce tobacco-related death and diseases [4–7].

There are many available mobile apps designed to support smokers in the process of tobacco cessation. However, most of these apps lack scientific evidence that prove their effectiveness [8]. In a recent systematic review [9] focused on analyzing available apps that support the process of smoking cessation, the authors found that only 6 out of 158 apps are supported by low-quality scientific evidence, 3 of which are currently available on smartphone markets, and only 2 of which are in the top 50 most popular apps for smoking cessation [9]. This creates two important issues: the limited number of scientifically validated apps and the unavailability of such apps for the general population.

In order to increase the quality and the availability of apps, experts suggest that it is necessary to develop an innovative framework capable of scientifically evaluating different properties of the apps, and then improve distribution channels to make those certified apps easier to find for the consumer. Moreover, certified apps will also be easier to use in health care environments, given that they should comply with a set of standards and regulations [10]. Official institutions also suggest the use of simple language to widen the dissemination of health-related information.

Our research team has been working since 2013 to evaluate the efficacy of a gamified mobile app with the goal of increasing the success rate of smoking cessation interventions in adults (individuals older than 18 years) that are already motivated to quit smoking. However, most of these apps are not supported by low-quality scientific evidence, 3 of which are currently available on smartphone markets, and only 2 of which are in the top 50 most popular apps for smoking cessation [9].

Methods

Study Design

The protocol of the study was previously published [11] and was executed in 2 phases. First, an interdisciplinary team composed of doctors, nurses, educators, designers, and computer engineers designed and implemented the Tobbstop mobile app, which combines gamification principles with the latest mobile technologies to create a novel experience designed to follow a tobacco withdrawal guide. Second, a cluster randomized clinical trial was conducted in order to evaluate the efficacy of the mobile app and the features included within.

Setting

We included participants from the primary health care regions of Tarragona and Terres de l’Ebre (Catalonia, Spain).

Recruitment of Primary Health Care Professionals

The participation of the primary health care professionals was voluntary. Health professionals—doctors and nurses—were informed of the goals and scope of the study, provided with the study protocol [11] and related documentation, and trained by the members of the research team to inform them of the details of the study. Individuals that participated in the study were recruited by the health professionals, provided that they met our inclusion criteria.

Participant Eligibility Criteria

Our inclusion criteria were (1) current smokers aged 18 years or older who smoked at least 10 cigarettes per day, (2) owners of an Android or iOS (Apple Inc) mobile phone compatible with our mobile app, and (c) smokers with a moderate-high motivation to quit smoking (Richmond test score ≥5) [12].

The exclusion criteria were (1) patients addicted to psychoactive substances other than tobacco, (2) patients with a psychotic disorder, (3) patients without a smartphone with the minimum hardware requirements necessary to run the app, and (4) participants who had a low motivation to quit smoking (Richmond test score <5).

If a participant did not meet the motivation criterion, health professionals provided information and measures to increase motivation and arranged a second session to try to recruit the participant into our study.

Random Allocation

The unit of randomization was the primary health care centers that participated in the recruiting. Each center enrolled was assigned randomly to one of the 2 groups (control or intervention). Centers were stratified according to rural or urban locations and the number of health professionals available in each of them. We ensured comparability between the 2 groups.

For the randomization, we used the software EpiData (version 3.0, EpiData Association).

Blinding

Given the nature of the intervention we could not prevent the participants and health professionals from knowing to which of the 2 groups they were assigned. However, in the data analysis phase, we blinded the data so no identification process could be carried out.

Intervention

During the recruitment process (June 2014–May 2016) professionals invited active smokers older than 18 years. The
participants that met the different selection criteria and were willing to participate were asked to sign a written consent form.

In the first visit (defined as visit 0), we collected demographic data, asked the participant to fill in the study questionnaire, and fixed the D-Day upon which they would initiate the process of smoking cessation. A few days before the D-Day, each participant had a second visit, where the smoker and the health professional discussed the plan to be followed throughout the process, in accordance with the Clinical Practice Guidelines of the Institut Català de la Salut—Catalan Health Institute [13]—and planned all the remaining monitoring visits.

Both groups received information on the standard guidelines of clinical practice. In the event that a participant did not attend a follow-up meeting, professionals called their phone to reintroduce them into the study. If they could not be reached, they were considered as relapsed as of that point.

**Description of the Mobile App**

The participants assigned to the intervention group (IG) received a numeric code to activate their access to the Tobbstop app and a detailed explanation of its basic features. This app is included in the Serious Games category, which are apps that include components from games designed to facilitate the achievement of the goals of the app, increasing user engagement and improving user experience. Tobbstop was designed with the goal of engaging the participants to use the app for at least the 3-month period that they were in the clinical study, and it included features to motivate them to use the app every day during this period. We also included features that covered Bartle’s taxonomy of player types (killers, socializers, achievers, and explorers) and that adapted to the different stages of the tobacco withdrawal process: start, euphoria, grief, normalization, and consolidation. All this work was performed with an interdisciplinary team of experts that included experts on tobacco withdrawal (in smokers and ex-smokers), computer engineers, graphic designers, game designers, educators, and health professionals (doctors and nurses). For more details about the app features, see Multimedia Appendix 1.

The clinical trial was designed to measure success 12 months after the start of the study, but the app was only designed to cover the first 3 months. Once participants completed the path through the island in the app, they could continue using the app and all the features provided without limitation.

**Control Group**

Participants assigned to the control group (CG) only followed the recommendations of the health professionals included in the study [13].

**Data Collection**

During the first visit, we collected all sociodemographic and anthropometric data, such as date of birth, gender, level of education, civil status, weight (kg), height (cm), and blood pressure (mmHg), as well as data on the presence of other pathologies, such as high blood pressure, chronic obstructive pulmonary disease, diabetes mellitus, stroke, neoplasia, acute myocardial infarction, dyslipidemia, angina pectoris, and intermittent claudication. We also collected data on tobacco consumption, including number of cigarettes smoked per day, usage of electronic cigarettes, age at which the individual started smoking, number of previous attempts to quit smoking, longest period without smoking, and presence of other smokers in the family. Nicotine dependence level was measured using the Fagerström test [14] and their motivation was measured using the Richmond test [12].

In each of the follow-up visits we measured weight and blood pressure and asked about tobacco consumption and the existence of abstinence syndrome. Tobacco abstinence was confirmed by the level of carbon monoxide (CO) in exhaled breath. We recorded any other treatments used for tobacco cessation (ie, pharmacological treatment).

If participants did not attend the follow-up interviews, we attempted to call them by phone. In some cases, when the participant was unable to attend the visit at the health center, we conducted the follow-up interview via phone in order to ensure the participant’s continuance with the study, but in such cases we were unable to collect all the measures (weight, height, CO-oximetry).

The follow-up period had the following end-points: (1) when the participant did not attend one of the visits and we were unable to contact them thereafter, (2) when the participant decided to quit the study, (3) when the participant started to smoke again, or (4) after 12 months without smoking.

**Computed Variables**

We computed the body mass index using height and weight and mean arterial pressure according to the formula [(diastolic arterial pressure x 2) + systolic arterial pressure] / 3.

**Sample Size**

To compute the sample size, which was randomized by the primary health care centers that participated, we multiplied the number of individuals required by the design effect. We accepted an α risk of 5% and a β risk of 20% in a bilateral contrast. We counted 222 participants in each group, detecting a difference of less than 5%, with measurements based on Epidat (version 3.1; Xunta de Galicia). In order to compute the design effect, we estimated an intraclass correlation coefficient in randomized clinical trials, which is usually lower than 0.05. The average size was 20 with a design effect of 1.36. Using all these values, we set the sample size to 604 participants, with 302 in each group.

**Statistical Analysis**

We performed intention-to-treat analysis to evaluate the comparability of the 2 groups.

Quantitative measures are described with mean and standard deviation if they have a normal distribution, or median and interquartile range if they do not. Qualitative variables are described with percentages and 95% confidence intervals. Baseline quantitative measures are compared using Student’s t test, while qualitative measures are compared using Pearson’s chi-square test.

The primary outcomes were continuous abstinence at 3 and 12 months.

http://mhealth.jmir.org/2020/6/e15951/
We evaluated the app efficacy using a protocol analysis, comparing the CG participants with the IG participants who used the app. We computed the crude and adjusted hazard ratios using a multilevel Cox regression with 2 effects (fixed and random). The fixed component included group assignment and, in the adjusted models, sociodemographic variables. The random component included assignment to a primary health care center. Data analysis was performed using R version 3.4.3 [15].

Ethical Aspects of the Study

The study, in its revised and updated version, was carried out following the Declaration of Helsinki principles and the Spanish Clinical Practice Guidelines. The study protocol was approved by the Clinical Research Ethics Committee at the Institut Universitari d'Investigació en Atenció Primària Jordi Gol. Data confidentiality was guaranteed by the Spanish law that regulated the protection of personal data at the time of the study, the Ley Orgánica de Protección de Datos de Carácter Personal (15/1999, December 13).

Results

Participant Characteristics

We recruited 773 participants for the study, of which 602 (77.9%) started the study on their D-Day. In Figure 1 we detail the flow of the participants within the study [16].

The participants who set their D-Day belonged to one of the 22 health care centers assigned to the IG, and they were treated by 67 health professionals (nurses and doctors). There were no significant differences in the basic characteristics of the participants prior to the intervention if we compare the centers in which they were recruited or the professionals who recruited them. The basic characteristics of the participants are detailed in Table 1. Even though there were not large differences between the 2 groups, we observed that the IG participants were younger than those in the CG (42.2 years vs 48.8 years; \( P < .001 \)), and they had a higher education level (\( P = .001 \)), lived alone (\( P = .03 \)), had a lower CO level (\( P = .001 \)), had lower dependence (\( P = .02 \)), had more family members who smoke (\( P = .03 \)), and had lower blood pressure (\( P = .003 \)) than participants in the CG. The CG used more pharmacological treatment for tobacco withdrawal (\( P < .001 \)).

When looking at the efficacy of the study by participant group, we observed that there were no significant statistical differences between the 2 groups, with slightly better results observed in the CG.

---

Figure 1. Flowchart of the participants included in the analysis.
Table 1. Basic characteristics of the Tobbstop study participants based on their assigned group.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control group (n=318)</th>
<th>Intervention group (n=284)</th>
<th>P value</th>
<th>Total (n=602)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, years (SD)</td>
<td>48.8 (11.0)</td>
<td>42.2 (10.2)</td>
<td>&lt;.001</td>
<td>45.7 (11.1)</td>
</tr>
<tr>
<td>Men (%)</td>
<td>155 (48.7)</td>
<td>120 (42.3)</td>
<td>.13</td>
<td>275 (45.7)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No studies or primary</td>
<td>126 (39.6)</td>
<td>70 (24.7)</td>
<td>.001</td>
<td>196 (32.6)</td>
</tr>
<tr>
<td>Secondary</td>
<td>129 (40.6)</td>
<td>143 (50.5)</td>
<td></td>
<td>272 (45.3)</td>
</tr>
<tr>
<td>University or higher</td>
<td>63 (19.8)</td>
<td>70 (24.7)</td>
<td></td>
<td>133 (22.1)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>55 (17.3)</td>
<td>72 (25.4)</td>
<td></td>
<td>127 (21.1)</td>
</tr>
<tr>
<td>With a partner</td>
<td>205 (64.5)</td>
<td>171 (60.2)</td>
<td></td>
<td>376 (62.5)</td>
</tr>
<tr>
<td>Widowed</td>
<td>15 (4.7)</td>
<td>5 (1.8)</td>
<td></td>
<td>20 (3.3)</td>
</tr>
<tr>
<td>Divorced</td>
<td>43 (13.5)</td>
<td>36 (12.7)</td>
<td></td>
<td>79 (13.1)</td>
</tr>
<tr>
<td>Body mass index (kg/m²), mean (SD)</td>
<td>27.4 (8.9)</td>
<td>26.9 (8.5)</td>
<td>.47</td>
<td>27.2 (8.7)</td>
</tr>
<tr>
<td>Mean arterial pressure (mm/Hg), mean (SD)</td>
<td>91.4 (10.6)</td>
<td>60.5 (10.7)</td>
<td>.30</td>
<td>91 (10.7)</td>
</tr>
<tr>
<td>CO₂-oximetry (ppm), mean (IQR)</td>
<td>16.8 (8.00-21.00)</td>
<td>13.6 (6.00-20.00)</td>
<td>.001</td>
<td>15.2 (6.75-20.00)</td>
</tr>
<tr>
<td>Electronic cigarette users, n (%)</td>
<td>49 (15.4)</td>
<td>54 (19)</td>
<td>.29</td>
<td>103 (17.1)</td>
</tr>
<tr>
<td><strong>Cigarettes per day, n (%)</strong></td>
<td></td>
<td></td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>0-10</td>
<td>60 (18.9)</td>
<td>73 (25.7)</td>
<td></td>
<td>133 (22.1)</td>
</tr>
<tr>
<td>11-20</td>
<td>183 (57.5)</td>
<td>154 (54.2)</td>
<td></td>
<td>337 (56.0)</td>
</tr>
<tr>
<td>21-30</td>
<td>60 (18.9)</td>
<td>38 (13.4)</td>
<td></td>
<td>98 (16.3)</td>
</tr>
<tr>
<td>31-40</td>
<td>12 (3.7)</td>
<td>15 (5.3)</td>
<td></td>
<td>27 (4.5)</td>
</tr>
<tr>
<td>&gt;40</td>
<td>3 (0.9)</td>
<td>4 (1.4)</td>
<td></td>
<td>7 (1.2)</td>
</tr>
<tr>
<td><strong>Fagerström test (dependence), n (%)</strong></td>
<td></td>
<td></td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>37 (11.6)</td>
<td>35 (12.3)</td>
<td></td>
<td>72 (12.0)</td>
</tr>
<tr>
<td>Moderate</td>
<td>217 (68.2)</td>
<td>165 (58.1)</td>
<td></td>
<td>382 (63.5)</td>
</tr>
<tr>
<td>Low</td>
<td>64 (20.1)</td>
<td>84 (29.6)</td>
<td></td>
<td>148 (24.6)</td>
</tr>
<tr>
<td>Number of previous attempts, mean (IQR)</td>
<td>1.90 (1.00-3.00)</td>
<td>2.61 (1.00-3.00)</td>
<td>.08</td>
<td>2.24 (1.00-3.00)</td>
</tr>
<tr>
<td>Maximum number of withdrawal months, mean (SD)</td>
<td>13.8 (31.1)</td>
<td>10.4 (20.9)</td>
<td>.12</td>
<td>12.2 (26.8)</td>
</tr>
<tr>
<td>Age of starting smoking, mean (SD)</td>
<td>17.0 (4.1)</td>
<td>16.7 (3.4)</td>
<td>.33</td>
<td>16.8 (3.8)</td>
</tr>
<tr>
<td>No smokers in the family, n (%)</td>
<td>154 (48.4)</td>
<td>111 (39.1)</td>
<td>.03</td>
<td>265 (44.0)</td>
</tr>
<tr>
<td>Use of a pharmacological treatment for tobacco cessation, n (%)</td>
<td>244 (76.7)</td>
<td>148 (52.1)</td>
<td>&lt;.001</td>
<td>392 (65.1)</td>
</tr>
<tr>
<td><strong>Pathologies, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>72 (22.6)</td>
<td>37 (13.0)</td>
<td>.003</td>
<td>109 (18.1)</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>24 (7.5)</td>
<td>11 (3.9)</td>
<td>.08</td>
<td>35 (5.8)</td>
</tr>
<tr>
<td>Type 2 diabetes mellitus</td>
<td>23 (7.2)</td>
<td>16 (5.6)</td>
<td>.53</td>
<td>39 (6.5)</td>
</tr>
<tr>
<td>Stroke</td>
<td>3 (0.9)</td>
<td>5 (1.8)</td>
<td>.49</td>
<td>8 (1.3)</td>
</tr>
<tr>
<td>Neoplasia</td>
<td>5 (1.6)</td>
<td>6 (2.1)</td>
<td>.85</td>
<td>11 (1.8)</td>
</tr>
<tr>
<td>Dyslipidemia</td>
<td>83 (26.1)</td>
<td>55 (19.4)</td>
<td>.06</td>
<td>138 (22.9)</td>
</tr>
<tr>
<td>Coronary heart disease</td>
<td>16 (5)</td>
<td>6 (2.1)</td>
<td>.09</td>
<td>22 (3.6)</td>
</tr>
<tr>
<td>3-month withdrawal, n (%)</td>
<td>101 (31.8)</td>
<td>85 (29.9)</td>
<td>.69</td>
<td>186 (30.9)</td>
</tr>
<tr>
<td>12-month withdrawal, n (%)</td>
<td>60 (18.9)</td>
<td>47 (16.5)</td>
<td>.53</td>
<td>107 (17.8)</td>
</tr>
</tbody>
</table>
App Usage

Out of the 284 participants enrolled in the IG, 97 (34.1%) did not use the app, never used the code to activate the app, or did not send any data on their usage from their mobile phone to our server.

We observed significant statistical differences between the participants who used the app and those who did not. Those who did not use the app had lower CO levels ($P=.03$) and smoked more cigarettes per day ($P=.04$). The IG participants who did not use the app indicated less pharmacological treatment than those who used the app (41/97, 42.3% vs 107/187, 57.2%; $P=.02$).

Finally, we observed that abstinence was significantly larger in the IG participants who used the app than in those who did not use the app at both 3 months (72/187, 38.5% vs 13/97, 13.4%; $P<.001$) and 12 months (39/187, 20.9% vs 8/97, 8.2%; $P=.01$; Figures 2 and 3).

Figure 2. Relation between entries and duration according to smoking cessation for 3 months.
Results According to the App Features Used

We consider the good users of the app (based on our design goals) to be those who used the app for at least 90 days and accessed the app at least 90 times. In accordance with these 2 conditions, we consider 26 of the 187 participants who used the app at least once to be good users. Among the main differences, we observed that the participants of the good users group smoked fewer cigarettes per day, had smaller dependencies according to the Fagerström test, and used less pharmacological treatment than those in the CG. The 3-month abstinence of the participants included in the good users group was 80.8% (21/26) compared with 31.7% (51/161) in the other users \((P<.001; \text{Figure 4})\). The 12-month abstinence rate for the good users was 57.7% (15/26) compared with 14.9% (24/161) for the other users of the app \((P<.001; \text{Figure 4})\).

We evaluated which app features were used the most by the IG. Those who succeeded were more active users of the app and logged in more than twice as much as the other participants. With the exception of the chat feature, which was used independently of whether the participant relapsed, those who did not relapse used the app features more than those who relapsed, as seen in Table 2. Out of the 187 participants who used the app, 38.5% (15/39) of the ex-smokers were considered good users, while only 7.4% (11/148) of those who relapsed were good users \((P<.001)\).
Figure 4. Kaplan-Meier curve GC (control group) vs GUG (good users group).

Table 2. List of a set of app metrics, classified according to the success rate at the end of the study (12 months from the day at which they stopped smoking) for participants of the intervention group who used the app (n=187).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Relapse (n=148)</th>
<th>Abstinence at 12 months (n=39)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacological treatment, n (%)</td>
<td>91 (61.5)</td>
<td>16 (41.0)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>App metrics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration within the app (days), mean (SD)</td>
<td>85.2 (94.8)</td>
<td>262 (220.0)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of different days connected, mean (SD)</td>
<td>31.7 (33.0)</td>
<td>67.4 (53.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Chat usage, n (%)</td>
<td>78 (52.7)</td>
<td>23 (59.0)</td>
<td>.60</td>
</tr>
<tr>
<td>Trivial maximum level, mean (SD)</td>
<td>12.7 (17.1)</td>
<td>18.3 (16.7)</td>
<td>.07</td>
</tr>
<tr>
<td>Trivial highest score, mean (SD)</td>
<td>612 (1011.0)</td>
<td>956 (989.0)</td>
<td>.06</td>
</tr>
<tr>
<td>Fruit game maximum level, mean (SD)</td>
<td>1.67 (1.3)</td>
<td>2.15 (1.3)</td>
<td>.04</td>
</tr>
<tr>
<td>Fruit game highest score, mean (SD)</td>
<td>1174 (2258.0)</td>
<td>2020 (2451.0)</td>
<td>.06</td>
</tr>
<tr>
<td>Number of challenges completed, mean (SD)</td>
<td>1.42 (3.1)</td>
<td>2.95 (4.5)</td>
<td>.05</td>
</tr>
<tr>
<td>Island sections completed, mean (SD)</td>
<td>6.26 (10.5)</td>
<td>12.5 (12.4)</td>
<td>.006</td>
</tr>
<tr>
<td>Number of times they consulted the information section, mean (SD)</td>
<td>3.61 (4.5)</td>
<td>6.44 (7.8)</td>
<td>.04</td>
</tr>
</tbody>
</table>

**Protocol Analysis of App Efficacy**

Observing that 34.1% (97/284) of the IG did not use the app, we evaluated app usage using a protocol analysis. When we selected the IG participants who had at least one activity record stored in the server and compared them with the CG participants, we did not find significant differences between the groups in terms of tobacco cessation at 3 months and 12 months (P=.26 and P=.94, respectively), as seen in Table 3.
We repeated the same analysis with the participants of the IG that we considered good users, taking into account that the size of the sample was small (n=26). We observed that those belonging to this group had a higher probability of being abstinent at 12 months (OR 7.20, 95% CI 2.14-24.20; P=.001) than the participants of the CG, as seen in Table 4.

Table 4. Adjusted association in smoking cessation at 3 months and 12 months between control group and good users.\(^a\)

<table>
<thead>
<tr>
<th>Clinical outcomes</th>
<th>Control(^b), n (%)</th>
<th>Good users(^c), n (%)</th>
<th>ICC(^d)</th>
<th>OR(^e) (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstinent at 3 months</td>
<td>101 (31.8)</td>
<td>21 (80.8)</td>
<td>0.000</td>
<td>9.88 (3.37-28.91)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Abstinent at 12 months</td>
<td>60 (18.9)</td>
<td>15 (57.7)</td>
<td>0.053</td>
<td>7.20 (2.14-24.20)</td>
<td>.001</td>
</tr>
</tbody>
</table>

\(^a\) Adjusted by clinic group, age, gender, body mass index, education, Fagerström test assessment, number of previous attempts to quit, smokers in the family, use of electronic cigarettes, and use of a pharmacological treatment for tobacco cessation.

\(^b\) n=318.

\(^c\) n=187.

\(^d\) ICC: intracluster correlation coefficient.

\(^e\) OR: odds ratio.

**Discussion**

**Principal Findings**

Our intervention trial, based on the use of a mobile app for smoking cessation in people older than 18 years who were motivated to quit smoking, included 773 participants who were monitored during their abstinence period for up to 12 months. The study demonstrated success in quitting smoking at 3 and 12 months among regular mobile app users compared with participants in the CG. Our study contributes to the literature on the design and evaluation of mobile health apps designed to help patients in the process of tobacco cessation [8-10,17], since most of the current research is based on the use of text messages to mobile phones [18].

Previous studies have shown that mobile phone app–based interventions may be useful tools for lifestyle interventions, such as weight loss [19], increased physical activity [20], or long-term condition management [21]. Moreover, a recent meta-analysis that included 5 studies that assess the effectiveness of smoking cessation using mobile apps alone to compare lower-intensity smoking cessations support found no evidence of a favorable effect of mobile apps in comparison with other types of interventions [22]. This meta-analysis also included one study that compared mobile app plus text messaging with a web-based intervention, which found evidence of a benefit of the app plus text messaging intervention [22]. Compared with the studies included in the meta-analysis, our study presents the longest follow-up (12 months) and the highest sample size.

We know that most attempts to quit smoking are not successful; according to the Centers for Disease Control and Prevention, only 12.2% of those who try to stop smoking remain abstinent [3]. For this reason, interventions that help young people and the general population stop smoking are needed. In our study, young people (aged 18 to 44 years) represent 57.7% (108/187) of those that used the app and 42.3% (11/26) of the good users. Mobile app usage seems more aligned with the lifestyle of the younger population, so this type of app has a double benefit in young people, since it reduces both smoking prevalence and comorbidity as patients grow older. Our mobile app includes some intervention components that have been successful in promoting smoking cessation, such as social compromise [23,24] and strategies to cope with abstinence syndrome and moments of craving [24,25].

Studies based on apps with this type of design should make sure that participants have a minimum level of digital skills with which to use the apps and that they are motivated to use them. As other researchers point out, it is important to study the usage patterns of the users in order to identify the features that are most helpful in the process of quitting smoking [26]. Researchers need to continue working on designs capable of identifying the key elements that help participants, as well as redesigning features (including new features) that can increase the efficacy of the app. Researchers should also consider how to personalize the intervention (and the available features within the app) to each of the population subgroups in order to maximize the engagement of the participants and, therefore, the probability of the success of the intervention. It is important to

Table 3. Adjusted association in smoking cessation at 3 months and 12 months between control group and app users.\(^a\)

<table>
<thead>
<tr>
<th>Clinical outcomes</th>
<th>Control(^b), n (%)</th>
<th>App users(^c), n (%)</th>
<th>ICC(^d)</th>
<th>OR(^e) (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstinent at 3 months</td>
<td>101 (31.8)</td>
<td>72 (38.5)</td>
<td>0.012</td>
<td>1.31 (0.82-2.09)</td>
<td>.26</td>
</tr>
<tr>
<td>Abstinent at 12 months</td>
<td>60 (18.9)</td>
<td>39 (20.9)</td>
<td>0.019</td>
<td>1.02 (0.58-1.79)</td>
<td>.94</td>
</tr>
</tbody>
</table>

\(^a\) Adjusted by clinic group, age, gender, body mass index, education, Fagerström test assessment, number of previous attempts to quit, smokers in the family, use of electronic cigarettes, and use of a pharmacological treatment for tobacco cessation.

\(^b\) n=318.

\(^c\) n=187.

\(^d\) ICC: intracluster correlation coefficient.

\(^e\) OR: odds ratio.
note that recent interventions that use an app for tobacco cessation are well received and viewed favorably by the participants [27], and it is vital that future apps take the needs of the users into account [28]. Mobile apps that use services hosted on online servers have an extra layer of complexity in complying with the technical and legal requirements of working with personal data.

In this initial Tobbstop study, users of the app appeared to exhibit patterns of participation and follow-up over time and demonstrated encouraging rates of tobacco cessation. Future research is warranted in order to evaluate the efficacy of Tobbstop in larger sample sizes.

We are currently working on adapting the app to other specific situations in which new features can improve its efficacy. In particular, we are designing a large-scale study of tobacco cessation among pregnant women. We have already performed a pilot study with 42 participants and have seen high success rates, even when the app was not designed for this specific group [29].

In our study, 34.1% (97/284) of the IG users never used the app, so we have to take this into account when making our analysis. This prevalence is similar to those obtained in previous studies [30]. Engagement and user retention are common and critical problems in mobile health. Previous studies have shown that more than two-thirds of people who downloaded a mobile phone app used it once and then stopped using it [31]. As a result, we have performed a sensitivity analysis between users and good users to obtain more accurate information about the efficacy of long-term usage of mobile phone apps in achieving smoking cessation. Our results have confirmed the hypothesis that long-term use of mobile apps improves the continuance of tobacco abstinence [32].

Despite the fact that the number of participant relapses were considered high, with 64.6% (183/284) of the IG and 73.3% (233/318) of the CG relapsing, our results have higher success rates than other studies that are also based on apps and have similar population samples [24,32]. In our study, we only enrolled participants with a high motivation to quit smoking. This parameter is unknown in other studies.

Conclusions
A mobile app to help the process of quitting smoking presents higher success rates than standard interventions, indicating the viability of conducting a randomized community trial based on smartphone technologies.

Acknowledgments
The authors would like to thank all who participated in the Tobbstop study, including all the health professionals; the Catalan Health Institute for publicizing our study; and the Institut Universitari d'Investigació en Atenció Primària Jordi Gol for all of the support we received.

The project received a research grant from the Carlos III Institute of Health, Ministry of Economy and Competitiveness (Spain). It was awarded by the 2012 call under Health Strategy Action 2013-2016 within the National Research Program oriented to Societal Challenges of the Technical, Scientific, and Innovation Research National Plan 2013-2016, with reference PI12/01977, cofunded by the European Regional Development Fund (European Union).

JD was funded by the Ministry of Economy and Competitiveness project (FIS2016-78904-C3-1-P), the Generalitat de Catalunya project (2017SGR-896), and the Universitat Rovira i Virgili project (2018PFR-B2-41).

Conflicts of Interest
None declared.

Multimedia Appendix 1
Main features and information about the Tobbstop application.
[DOCX File, 862 KB - mhealth_v8i6e15951_app1.docx ]

Multimedia Appendix 2
CONSORT checklist.
[PDF File (Adobe PDF File), 195 KB - mhealth_v8i6e15951_app2.pdf ]

References


16. CONSORT. Welcome to the CONSORT Website URL: http://www.consort-statement.org/ [accessed 2020-05-31]


Designing for the Co-Use of Consumer Health Technology in Self-Management of Adolescent Overweight and Obesity: Mixed Methods Qualitative Study

Cynthia M LeRouge\textsuperscript{1,}*, PhD; Hyeyoung Hah\textsuperscript{1,}*, PhD; Gloria J Deckard\textsuperscript{1}, PhD; Haoqiang Jiang\textsuperscript{2,}*, PhD

\textsuperscript{1}Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States
\textsuperscript{2}College of Informatics, Northern Kentucky University, Highland Heights, KY, United States
* these authors contributed equally

Corresponding Author:
Cynthia M LeRouge, PhD
Department of Information Systems and Business Analytics
Florida International University
11200 SW 8th Street
Miami, FL, 33199
United States
Phone: 1 3147090624
Fax: 1 3053484709
Email: clerouge@fiu.edu

Abstract

Background: Overweight and obesity in adolescents has reached epidemic proportions in the United States. Consumer health technology (CHT) can serve as a behavioral and social support tool for the management of overweight in adolescence. Recognizing CHT as a social support tool during design enables input from multiple stakeholders who engage in shared co-use to reinforce and empower adolescents in their self-management efforts.

Objective: This study aimed to explore design requirements and enabling factors for the use of CHT as a social support tool for patients (as primary users) and parents and health care providers (as co-users). Our model incorporates key components of the unified theory of acceptance and use of technology (UTAUT) within the framework of the obesity care model (OCM) by recognizing patient self-management as the central process with the influence of their care support network on CHT use and outcomes.

Methods: This study was part of a larger two-staged usability study combining focus group, semistructured interviews, and usability walkthroughs of CHT mockups from adolescents (BMI in the 85th-99th percentile range), parents, and physicians. In phase 1, 48 adolescents between the ages of 12 and 17 years, 10 of their parents, and 6 health care providers participated in identifying design requirements and enabling factors for the use of a potential CHT. In phase 2, 70 adolescents and 10 health care providers evaluated the CHT mockups and indicated enabling factors and willingness to use the proposed CHT.

Results: Our qualitative analysis identified adolescents’ intention for the use of CHT in alignment with UTAUT elements of performance expectancy, effort expectancy, and facilitating conditions. Our reconceptualization of social influence identified the expectations and envisioned roles of parents and health care providers as co-users and influencing factors on the co-use of CHT in managing overweight in adolescence. Parents were expected to monitor, to provide guidance and motivation, and to suggest modifications in daily habits, for example, recipes and meals, whereas health care providers were expected to encourage and monitor progress in a clinical setting. These expected roles and co-use patterns were congruent among all 3 stakeholders; the co-use of CHT was desired to be minimally invasive for parents and health care providers and controlled by the adolescents.

Conclusions: Our study integrates and extends the perspectives of 2 seminal models to explore design features and social influence roles for the successful user-centered design of CHT for weight self-management in adolescents. Although the co-users (ie, adolescents, parents, health care providers) suggested differing features consistent with their roles, role definitions were congruent. All users recognized the adolescent as the primary user with differential, supportive use from parents and health care providers. This multistakeholder approach can guide successful CHT design that reinforces the collective perspective of self-management.
Introduction

Background

The prevalence of childhood overweight and obesity has steadily increased over the past 4 decades in the United States and remains to be high (approximately 1 in 5) [1]. Categorized as weight higher than that considered healthy for a given height, the Centers for Disease Control and Prevention suggest a relative BMI greater than or equal to the 85th percentile as overweight and BMI equal to or greater than the 95th percentile as obese (throughout this paper, to facilitate readability, we will utilize the term overweight as a descriptor for both conditions when the distinction between the 2 conditions is not needed for clarity) [2]. These youths suffer both short- and long-term physical and mental health outcomes [3,4]. Overweight is linked to higher risks for many chronic diseases, including cardiovascular disease, asthma, and type 2 diabetes, which can result in a shorter lifespan [4-7]. Numerous studies have found a strong association between weight status and depression and other mental illnesses [8-10]. In addition, overweight youth are at a greater risk of having obesity and associated health risks as adults [11].

With an already high prevalence of adult obesity in the United States, prevention is a necessary step in reducing both short- and long-term health outcomes, decreasing health care expenditures, and improving quality of life. Recognized as a chronic condition, the etiology of overweight is complex with no definitive cause, although lifestyle behaviors related to food habits, eating patterns, and physical activity are known determinants. Given that life-long habits developed in early childhood and adolescence underlie the foundation for habits in adulthood, early preventative health management is pivotal to addressing the overweight epidemic [12].

Our study recognizes consumer health technology (CHT) as a social tool that enables multistakeholders in the care support network to engage in overweight management efforts by adolescents, thereby reinforcing self-management capabilities. Although previous studies have emphasized the ever-increasing role of patients in the care process, CHT design in the context of a support network has been underexplored [13,14].

Conceptual Background

Consumer Health Technology Role in Support

CHT can play an important role in the self-management of patients with chronic diseases, such as overweight and obesity. Ranging from mobile apps to patient portals, CHTs are broadly defined as information and communication technologies used by consumers for health purposes [15]. The benefits of CHT have been widely discussed with respect to its capability of care transformation, streamlining communication between patients and clinicians, and empowering patients for health decision making. Previous studies highlight that CHT can be integral to enabling (1) a feedback chain (which supports the user in discerning behavioral causes producing either desired or undesired outcomes), (2) information exchange among the user and support network, (3) data management, (4) care planning, and (5) social support from family members [12,13,16-18]. Thus, CHT can provide a means to customize a care regimen by connecting patients to their care team [19-21]. CHT can also function as a decision support tool for both patients and health care providers [21-23]. Tailoring CHT to the preferences and context, for example, the care support network as co-users, can promote the CHT’s potential to reinforce a healthy lifestyle in overweight adolescents [24].

To address a tailored, user-centered approach to deploy CHT, it is important to recognize that overweight management in adolescents is not an isolated health task. Health care self-management in adolescents comprises socially collaborative efforts from adolescents, family members (particularly parents), and care health care providers. Parental influence may be particularly salient because of the sharing of living space, meals, physical activities, and sleeping habits. Overweight adolescents also need guidance and advice from health care providers who can design an individualized care plan for the eating patterns and levels of physical activities of adolescents. Given their integral yet differential roles for influence and support, parents and health care providers may influence how adolescents use CHT.

Caregiver Role in Support

Effective prevention and care of chronic illness require an organized delivery system linked with multiple supportive community resources [25]. Such support systems are not generally recognized nor the needs of those with chronic illness adequately met in a traditional primary, acute health care system [26]. The chronic care model (CCM) addresses these complementary systems (Figure 1) and provides an approach to high-quality chronic care characterized by productive interactions between health system health care providers and patients that consistently provide assessments, support for self-management, optimization of therapy, and follow-up (Figure 1) [27].

Development of the CCM was based on evidence from system interventions to improve care for various chronically ill populations, and subsequent implementation generally focused on changes within care delivery organizations [28,29]. To provide more focus on the environments outside the clinic or health system walls, the obesity care model (OCM) modifies the original CCM framework to place equal emphasis on the community and the medical system (Figure 2) [30]. The OCM places patient and family self-management as the central process to improve weight management. As shown in Figure 2, the role of health care health care providers to assist in self-management...
includes information systems and decision support, which may be implemented through CHT [31].

The CCM and OCM provide perspectives and structures to investigate the multilateral roles adolescents, parents, the community, and health care providers play in managing the adolescent’s weight. These critical players may also play an important role in the success of CHT as secondary users or co-users. CHT can streamline shared care tasks among patients, parents, and health care providers. In translating OCM or CCM concepts to CHT, members of the adolescent’s care support network (eg, parents and health care providers) may act as co-users by noting goal accomplishments or sharing information (eg, motivating messages); health care providers can provide feedback and support as well as contribute to the care plan.

**Figure 1.** Chronic care model.

**Figure 2.** Obesity care model.

**Recognizing the Co-User Role in Adoption and Use**

The phenomena of a care support network in relation to CHT and co-users of technology can be theoretically linked to the notion of social influence referenced in the unified theory of acceptance and use of technology (UTAUT) [32]. The UTAUT identifies 4 determinants of attitude and intent to use information technology: (1) perceived usefulness (performance expectancy), (2) effort expectancy (ease of use), (3) social influence (significant others’ evaluation of technology use), and (4) facilitating conditions (supporting infrastructure for technology use). These UTAUT determinants help us explain technology adoption behaviors by health care consumers and health care providers [33-38]. Understanding, contextualizing, and reacting to the precursors of intent to use by the primary technology user (ie, adolescent) is a critical foundation for the design and adoption of CHT. Extending this understanding to supportive co-users (ie, secondary users) in alignment with OCM or CCM concepts can improve CHT design and increase the understanding of the contribution of the co-user roles.

The UTAUT social influence antecedent recognizes that an individual’s intention to use technology may adjust or modify their technology adoption behavior (a concept comparable with the environmental influences found in the OCM). The care support environments referenced in the OCM may provide social influence (particularly through indirect or direct engagement with the CHT that may impact the primary user’s opinions or behaviors toward using the CHT and self-management) to further explore the possibilities of CHT to assist adolescents’ self-management efforts (Figure 3). Research shows that social influence impacts dietary behavior and adolescents’ readiness toward their chronic care health [39,40]. Unfortunately, there is a lack of literature that integrates the co-user’s perspectives and expectations regarding the intention to adopt CHT for adolescent overweight management.
Goal of This Study

Our research model presented in Figure 4 utilizes the UTAUT to identify the impact of social influence on the co-use intention for adolescents, parents, and health care providers. The model also incorporates the OCM by recognizing the use, support, and influence of the primary CHT user (the adolescent) by parents and health care providers. That is, a patient’s CHT use behavior can be influenced not only by the extent to which an individual believes how others view his or her use of technology but also by the way others in the patient’s support network and secondary users assess the technology and the potential benefit of the patient’s technology utilization. Our research model simultaneously incorporates the factors that shape a primary user’s CHT adoption and that of secondary users.

In this model, we reconceptualize the social influence construct that can connect the adoption of 3 distinct users (patient, provider, and caregivers, including parents) in the same nomological care network. The social influence construct takes

Figure 3. Design perspective of primary and co-user roles in consumer health technology.

Figure 4. Intention to co-use model. CHT: consumer health technology.
the primary users’ perception of the use and role of co-users into account when explaining the primary users’ information technology adoption intention. We specify social influence for our context of interest as the way an adolescent perceives the role of others in the OCM (eg, health care providers, community including peers, and family) for their use in CHT apps.

We apply this model to a qualitative study of CHT design in the context of a support network for overweight and obesity in adolescents by first contextualizing the UTAUT constructs of performance expectancy, effort expectancy, and facilitating conditions as antecedents of use of CHT for adolescent overweight and obesity by the primary user (adolescent). We will then reconceptualize social influence in relation to the co-use of CHT in the social cycle of adolescent-parent-health care providers supporting the management of adolescent overweight and obesity. More specifically, we examine how multiple users perceive their shared care tasks that can further lead to their envisaged co-use intention. This study’s unique attention to each user’s perception and expected design requirements for CHT prototypes provides an in-depth understanding of feature-based CHT design linking to intention to use CHT from multiusers.

### Methods

#### Methodology

This qualitative study integrates the UTAUT framework within the OCM context for the management of overweight in adolescents. Through this integration of support and use concepts, we leverage a user-centered design methodology to examine users’ adoption of CHT for overweight management in adolescents.

This paper is part of a more expansive user-centered design research project that explored questions and issues related to requirements analysis, human-computer interaction, prototyping, usability, technology evaluation, and technology adoption by overweight adolescents, their caregivers, and their health care providers regarding technology to support self-management and caregiver connections. This design project included 2 phases: (1) analysis of needs, preferences, and human-computer interaction in social context and (2) design, prototyping, and usability. We received an institutional review board approval from Saint Louis University. Here, we focus on aspects of the methods, particularly related to the research intentions of this study (further details of the global study are reported elsewhere) [41].

To demonstrate rigor with respect to study validity, we carefully considered the research methods used (ascribed to user-centered design approaches), the participants (sources of data), and the number of participants. For qualitative studies, the sample size may be considered a matter of judgment and experience in evaluating the quality of the information collected against the uses to which it will be put, the particular research method, and purposeful sampling strategy that focuses on individuals with the experience to provide appropriate insights for the targeted population [42]. A study of sample size for qualitative research by Faulkner found that the exact numbers required for usability tests were not as important as the participants representing the target population [43]. For each stage and with each method of data collection, selection of our participants was purposeful to include individuals with direct knowledge of the points of inquiry to accurately reflect the targeted user population (ie, overweight or obese adolescents) [42]. We balanced the purposeful determination of appropriate informants (to share insights on the adolescent as primary use and parents and health care providers as secondary users) with heuristics on sample size and the overall need to achieve saturation in identifying general goals for data collection and points of saturation. Saturation is defined as the point at which additional data collection no longer generates a new understanding [44]. For qualitative research in general, a recent review showed that the median sample size was between 28 and 31 to achieve data saturation [45]. In total, this study included 143 individuals, far exceeding this norm. However, this number should be interpreted with some caveats. The sample size in qualitative research may refer to the numbers of persons (say for individual interviews or individual thinking out loud activities as a usability exercise) but also to the numbers of interviews (such as a focus group), given the different units of analysis [42]. We provide further details, as appropriate, in each phase to further showcase rigor.

#### Phase 1

The goal of phase 1, with respect to the scope of this study, was to detect desired application areas to assist with self-management and to identify design and functional requirements as a means to inform design details to meet the adolescents’ performance and effort expectancy (UTAUT model). It was also important to determine the context of self-management (facilitating conditions from UTAUT and OCM) for these adolescents. In addition, this phase began to explore the elements of productive interactions (CCM model) among adolescents, their families, and health care providers to their self-management.

Phase 1 techniques included 10 one-hour user-driven focus groups with adolescents participating in 2 different sessions in an accredited weight self-management program in the Midwest, United States. The weight loss camp allowed the researchers to access individuals with relevant experience and appropriate insights on the usability of apps for weight self-management by adolescents. The summer camp program (specifically, Camp Jump Start, a recognized adolescent weight loss and healthy living summer camp program, grounded in evidence-based techniques and led by a health care provider) attracts clients from across the United States. The study participants included a diverse group of black, white, and Hispanic adolescents, aged between 12 and 17 years, from various socioeconomic backgrounds. Both males and females were included in the study. The inclusion criteria were based on age, computer use, and BMI (participants were in the 85th-99th percentile range).

Each focus group had 3 to 7 participants, and the total number of participants for the focus groups was 48. Quantitatively, it is recommended to have 3 to 5 groups per project to reach saturation [46]. Further, the general recommendation is that each group should have a minimum of either 3 to 4 and a maximum of 12 participants [46,47].
Then, 15 semistructured interviews with the parents of adolescents participating in the summer camp and 6 semistructured interviews with pediatric physicians of these children were conducted. Selected focus group and semistructured interview questions were developed with consideration of the UTAUT, CCM, and OCM factors and CCHT with other key users (e.g., parents and health care providers). Underpinning co-use, as part of social influence, are the first 3 constructs that serve as antecedents to intention to use (performance expectancy, effort expectancy, facilitating conditions, and social influence) [32]. The first 3 constructs are closely linked to the environment and supportive concepts of the OCM. In alignment with our model, we leverage constructs from the UTAUT and OCM or CCM models to afford more detailed levels of explanation of each of the constructs in the integrated UTAUT and OCM or CCM models to further understand the nature of each of these constructs in the context of the study. Third, another research team member independently reviewed all codes (100% code review) to determine the propriety of the assigned codes to the constructs in interest, noting potential issues and questions to the coding process to iteratively discuss initial coding, refine coding categories, and reach consensus regarding identified subthemes [50]. Reconciling issues were minor and primarily consisted of agreeing on combining child codes, appropriate titles for code names, and addressing coder thoughts on whether a new code was required. Data analysis was complete after the team reached a consensus that all quotes were appropriately coded, and the resulting themes under each construct were stable [51].

### Results

#### Overview

Our investigation underscores 2 different aspects of adolescents’ CHT use: primary use and co-use with other members of their care support team. That is, adolescents act both as primary users and co-users, with parents and health care providers participating as secondary users in supportive roles consistent with the environment and supportive concepts of the OCM. In alignment with our model, we leverage constructs from the UTAUT model that serve as antecedents to intention to use (performance expectancy, effort expectancy, facilitating conditions, and social influence) [32]. The first 3 constructs are closely linked to the primary (sole) use of CHT, whereas social influence demonstrates the extent to which adolescents desire co-use of CHT with other key users (e.g., parents and health care providers). Underpinning co-use, as part of social influence, are
the adolescent’s expectations of desirable inputs and feedback from co-users.

In the following section, we first link the UTAUT constructs of performance expectancy, effort expectancy, and facilitating conditions to adolescents’ perceptions of the primary use of CHT. Then, we examine the perceptions of the role of co-users as social influence that may alter adolescents’ intention to act as primary users and lead to the co-use of CHT with parents and health care providers.

Performance Expectancy

In our research context, performance expectancy is viewed as the degree to which an adolescent believes that the availability and use of CHT functions and features may help him or her with weight management. Our results indicate that adolescents anticipate that a successful CHT app would promote, reinforce, or extend their motivation to combat obesity, try healthy food alternatives, and exhibit healthy behaviors. Viewing CHT as motivational support to their busy schedules between school, extracurricular activities, and social obligations, they identified a list of key features related to behavioral awareness and guidance and education on healthy living. In essence, features that would provide them with tools that underpin and promote self-management.

CHT features identified by the adolescents related to behavioral awareness and guidance included (1) timely reminders and automated notifications to log in and complete information (eg, food diary entries), (2) prompts to take certain actions (complete a physical activity), and (3) educational information on healthy food and living choices. To assist in learning healthy behaviors and altering lifestyle habits, adolescents suggested CHT features that would (1) set personal goals, (2) track their daily behaviors (eg, caloric intake and calorie burning), and (3) provide rewards (eg, awards, coupons, or discounts for healthy foods at stores or restaurants). They also expressed a desire to keep all of the tracking information and materials in one location on a computer or smartphone. To keep track of their progress, they were interested in visualizing graphs of their weight loss efforts.

Effort Expectancy

We regard effort expectancy as the amount of effort adolescents are willing to exert to use CHT in their self-management of overweight. Our interviews revealed that adolescents expect ease of use of the system interface as one of the most important factors for the adoption of CHT. Adolescents mentioned that minimal effort was made possible by an uncomplicated interface to easily find and add their own health information. To reduce writing or typing efforts, features such as drag-and-drop capabilities, drop-down menus, and search function were mentioned as means to ease efforts to find, add, and edit personal information. Furthermore, adolescents desired the system to display their progress and history readily without much effort.

In addition, they wanted CHT apps to be flexible, for instance, to allow infrequent data entry, smooth integration among its modules, and connections to other external apps. The adolescents also desired the capability to edit app components such as recipes so that they could build or edit recipes to fit their own tastes.

Simplicity in use was particularly important for smartphone interfaces, as some adolescents indicated that they would like some level of access on the go via smartphones, given the smaller screen size of the phone.

Finally, the adolescents suggested an entertainment component, such as games and avatars, to be featured in the CHT app. They indicated that game features, along with the ability to share game progress, would motivate them to return to the app and reduce their perceived effort to use.

Facilitating Conditions

Facilitating conditions reference a belief by the adolescents that certain technical and contextual conditions support their use of the CHT app. The roots of this UTAUT construct include perceived behavioral control, facilitating environmental conditions, and compatibility. Our results show that adolescents deemed compatibility as a major factor impacting their use of the CHT. Adolescents indicated that access to the CHT app from multiple platforms or devices such as phones, tablets, and computers would enable them to track progress more easily. The most preferred platform was smartphones, as most adolescents stated that being able to track while on the go would support their frequent usage of the app. Lack of advertisements was also mentioned as a facilitating condition for using the apps.

The adolescents identified a range of conditions that could be viewed as nonfacilitating, particularly technical issues that could prevent their use of the CHT. These potential barriers included the cost of the app, password memorization, and system crashes. In addition, connectivity issues associated with Wi-Fi or the internet could block CHT use.

Overall, overweight adolescents view CHT apps as motivational and capable of altering behavioral patterns. Our analysis of primary user perceptions for intent to use a CHT identify key features and factors compatible with their day-to-day use for self-management. Acceptance and use of the CHT depend on the incorporation of features noted above into the design of the app.

We now turn to our attention to perceptions of the primary user (overweight adolescents) and the supporting community of parents and health care providers regarding the social influence and co-use of CHT.

Social Influence as Co-Use of Consumer Health Technology

In our research context, we integrate the UTAUT concept of social influence that focuses on individuals’ perceptions that the people close to them believe that they should try using the technology with the OCM, which considers patient and family engagement and empowerment. Our conceptualization examines the degree to which an adolescent perceives the role of others in the OCM, for example, parents and health care providers, and their co-use of CHT in their self-management of overweight. We also explore the perceptions of co-use by the health care providers and parents (the OCM support network) of these adolescents regarding their role and use of the CHT.

http://mhealth.jmir.org/2020/6/e18391/
**Social Influence and Co-Use: Adolescents’ View**

Adolescents identified key participants in their community of support and social influence as family, including parents and siblings, health care providers and peers or friends, and nominated parents and health care providers as co-users of the CHT app in their weight loss management program. They viewed the role of social influence to include acknowledging challenges, setting up motivational tasks, respecting inputs, and sharing achievements.

By sharing with others through co-use of CHT, the adolescents thought they could discuss difficulties and achievements, mindfully monitor their progress, and receive encouragement from their support community in a timely fashion. An exemplary quote summarizing these thoughts states:

*You could talk to other people about your difficulties and they could give you tips and advice and just help each other out.*

Simultaneously, however, adolescents draw a clear line regarding the boundaries of co-use in terms of data, information, and management. Self-management should be guarded so as not to give up control. Others participating in cousage should not be able to modify the information or alter the adolescents’ weight loss process, for example:

*I wouldn’t want them to like go in and mess everything up.*

**Adolescents’ View of Co-Use by Family Members**

Although adolescents acknowledged multiple family members as potential co-users, they clearly identified parents as one of the major sources of social influence. They indicated that parents could provide support by reviewing progress, providing encouragement, cocreating information content, for example, recipes, and encouraging family interaction. Adolescents also mentioned creating a competition between family members—each setting up weekly goals and seeing who does best.

Generally, parents as co-users of the app would demonstrate engagement while empowering adolescents’ self-management and drive toward achieving their goals, providing motivational support, and adding accountability by reviewing their progress.

**Adolescents’ View of Co-Use by Health Care Providers**

The adolescents viewed the role of health care providers’ social influence as creating the obesity care plan and encouraging through monitoring progress, providing feedback, and acknowledging achievements. Having the guidance and support of their health care provider outside of visits to physicians provides adolescents with confidence during different steps of their behavioral modification process.

Our results show that adolescents confirm the importance of and need for social influence and a supportive community from parents and health care providers to provide guidance and assistance with their self-management of overweight. However, this influence has clear boundaries and should be supportive and empowering, not interfering. Defining who can help them and how others can help them within their desired social influence is instrumental to a successfully designed CHT that all users (primary and co-users) will engage with as intended.

**Social Influence and Co-Use Intention: Parents and Health Care Providers**

The perceptions of co-users regarding their role and use of CHT are a critical, yet often overlooked, component for both design and use. Here, we report the perceptions of 2 key co-users, parents and health care providers, whose social influence on overweight adolescents is an essential component of our integrated conceptualization of the concept.

**Parents’ View of Co-Use**

Our interviews with parents revealed their willingness to support their children and to participate as co-users of CHT designed to assist their self-management of overweight. Parents identified forms of social influence and support and desirable characteristics of the CHT.

First, parents clearly recognized the adolescent as the primary user of the CHT and themselves as a secondary user. Although wanting to provide support in a number of ways, they acknowledged the need to respect the privacy of their children. Parents expressed these perceptions through statements such as the following:

*I can see myself monitoring, as a kind of oversight... But if you get too involved, then it’s no longer interesting to them... they feel like it’s invading their privacy.*

To assure the adolescents of their privacy, parents indicated that having their own log-in was important whereas tying specific aspects of the CHT together for input and support:

*...you’re probably more successful together. You have someone to talk with and give you support.*

Parents defined their support role as one to provide guidance, shared accountability, and motivation:

*I could see myself being more involved, you know, like a participative role where maybe he’s coming up with something and I’m giving him feedback.*

Yet doing so, only within the areas sanctioned by the adolescent as indicated by this sentiment:

*I think there’s a part of me who would really like to peek over her shoulder and see everything she’s doing, but I also feel like she deserves her privacy, and she deserves that respect if she wouldn’t want to share it... I think I would really want to go just based on her comfort level.*

The value of the CHT itself for motivation and guidance was equally recognized by parents:

*I think it could be helpful if it could keep her motivated.*

Regarding features that would support their co-use, the parents suggested the technology to be general enough to monitor the progress of their children and allow them to provide timely advice or input without being invasive. One thought expressed
was to provide icons such as a “yellow happy face” to provide encouragement.

Parents are clearly willing to provide their children with support to utilize CHT tools in an everyday setting and to provide support through the co-use of the CHT. Interestingly, parents distinguished themselves as secondary users in supportive roles that encouraged their children’s self-management as primary users of the CHT.

**Health Care Providers’ View of Co-Use**

Health care providers perceived themselves as a source of encouragement and support in adolescents’ obesity management. Health care providers acknowledged their role in providing medical system support to go beyond office visits and referenced the need for a means to acknowledge adolescents’ progress and provide feedback on a regular basis. Given that health care providers manage a number of adolescents with obesity, they readily acknowledged that their ability to communicate and extend this support between office visits could be streamlined by the co-use of CHT.

Due to daily time constraints in physicians’ offices, health care providers expressed their strong interest in giving encouragement by using the automatic CHT system features to quickly send their rewards, notes, or feedback that “wouldn’t be hard and wouldn’t take a lot of time.” Minimizing time and effort is critical for health care providers’ co-use of the CHT application.

In addition, health care providers wanted the CHT to track adolescents’ progress and provide a quick summary report showing how their patients are progressing toward their individual goals. The health care providers also indicated that they would value an alerting feature that noted both achievement and lack of progress or lack of data input. Alerts would allow a real-time response and signal the provider to contribute appropriate, timely feedback through the CHT to the adolescent.

Overall, health care providers are inclined to encourage adolescents’ primary usage of the CHT and want the ability to reward their achievements or encourage progress through efficient co-use capabilities. As health care providers’ time is limited and they may manage multiple primary users (patients), effective co-use relies on automatic features to monitor their patients’ daily progress, provide weekly reports, and efficient means for providing support to adolescents for obesity management.

**Discussion**

**Summary of Insights**

This paper examines the factors influencing the intent to use CHT in the management of overweight by adolescents and the role of social influence from the perspective of adolescents, parents, and health care providers. Social influence cannot only impact the use of CHT as conceptualized by the UTAUT model but also provide meaningful engagement of parents and health care providers in a way that empowers adolescents’ self-management of overweight.

Through the inclusion of both primary users (adolescents) and co-users (parents and health care providers) in our study, we can contrast perspectives to assess the congruence in the defining primary user and co-user roles for CHT in the self-management of adolescent overweight. Recognition of the adolescent as the primary CHT user is consistent for all 3 contributors. Both parents and health care providers view their roles as secondary or co-users and are responsible for providing information, feedback, encouragement, and support. Differences in the roles of these secondary users are apparent, as health care providers are part of the medical system, whereas parents are part of the environment of family and community support. Overall, health care providers were viewed in the expert role providing plans and directives along with motivational feedback, whereas parents were seen as critical in supporting, encouraging, and motivating adolescents in the process of self-management. Rather than interfering, the roles of both co-users were directed toward empowering adolescents. The medical system is more independent from the primary user and assumes more oversight of the adolescent’s progress, albeit through automated CHT features. Parents are more interdependent; however, shared access is controlled by adolescents in recognition of the need for privacy and respect.

Our findings point to the importance of the visibility and transparency of co-use in the realm of CHT. In our study, adolescents would choose the activities and inputs visible on the CHT to specific co-users, possibly with some sensitive information hidden from parents but available to health care providers. However, the activities of both parents and health care providers as co-users are to be visible to their children to create a noninvasive, shared sense of accountability for obesity management.

By explicitly looking for role congruence regarding social influence from parents and health care providers, we found that as co-users, both parents and health care providers are willing to engage in the process and fulfill their role to empower the primary user consistent with the OCM. These findings correspond to the existing studies that highlight the needs of including parents and physicians in the process of mobile health (mHealth) app design and implementation to promote the use of mHealth technologies [52-57].

From an adolescent’s perspective of CHT use in overweight self-management, our results establish that adolescents are willing to (1) use CHT features that promote behavioral awareness and guidance, (2) input and monitor progress toward weight loss goals across platforms, and (3) allow motivational inputs and encouragement (social influence and support) from parents and health care providers as co-users within defined areas that respect their privacy. As co-users of the CHT, parents and health care providers recognize that adolescents want respect for their privacy whereas desiring feedback to motivate and inspire. The provider sought independent and automated feedback features whereas parents desired to project shared accountability, pulling together to motivate achievement of goals and encouraging progress. We invite future research to explore the need and benefits of sustained use for primary users and co-users.
When designing CHT for obesity management, incorporating the perspectives of the OCM into UTAUT can provide a broader, more comprehensive view of social influence—one that provides input from patients as primary users, parents as part of the supportive environment, and health care providers as part of the medical system. Responding to each of these groups, their concern for CHT features and the role they play can improve successful design and adoption.

**Contributions of the Study**

The contributions of this paper are three-fold. First, we investigate the construct of the social influence of the UTAUT framework through the lens of the OCM and incorporate perspectives of 3 users. We expand the social influence construct by exploring the primary user’s specific expectations of secondary users, namely family members (particularly parents) and health care providers, for social support as an antecedent to use CHT to manage adolescent overweight. Our findings point to the dynamic nature of social influence among adolescents, parents, and health care providers. Perceptions of the varying roles demonstrate congruence among the users; this common recognition of adolescents as independent, primary users may empower and encourage their ongoing use of CHT in the self-management process.

Second, we propose the role of CHT co-use for overweight to be tailored to the OCM framework that extends the intention to use construct to recognize CHT co-use (intended behaviors for use) by family members (particularly parents) and health care providers as secondary users that simultaneously influence the adoption of CHT for managing overweight in adolescents. Acknowledging the importance of designing a user interface that is appealing to the users and responds to their needs is central in developing user-centered CHT systems [51-57]. This study’s approach to the design of CHT allows for simultaneous use in the management of adolescent obesity across 3 system users. We found that in the OCM context, the co-use of CHT may support parents and health care providers as supportive social influencers in adolescents’ overweight management process while empowering adolescents in the leading role. To the best of our knowledge, this is one of the earliest studies to consider co-use in a chronic care partnership with adolescents. Such a co-use care partnership should be extended with a series of further studies.

Finally, our user-centered design approach identifies envisioned features that directly influence intention and use for 3 co-users. The design requirements revealed from all 3 user groups can further lay a groundwork to tailor and contextualize the features of CHT systems in the chronic care management context.

With the growing role of CHT in the treatment of chronic care, our study has many practical implications for enhancing active co-use from patients, parents, and health care providers. Our study suggests that CHT vendors provide a user-centered design for multistakeholders to manage shared health care tasks for active participation and effectiveness in self-management. To better promote a broader range of CHT usability, vendors and developers need to relate to co-users and accommodate the intended users’ needs and requirements.

Our study may also inform health system health care providers who need to focus on the impacts of coengagement from parents and care health care providers in special care task contexts. Parents generally influence the lifestyle habits and decision making of children and adolescents. Therefore, understanding the co-use of CHT allows for articulation of the roles of parents (or caregivers) and care health care providers to provide tailored CCM and OCM plans for various care contexts.

**Limitations**

As with most qualitative studies, a generalization of the results must be approached with some caution. Although the study included a diverse set of adolescent participants, the study involved a limited sample of adolescents between 12 and 17 years. Our findings may not be applicable to children younger than 12 years of age or older than 17 years. First, the study also included a limited sample of parents of these adolescents and health care providers serving the adolescent population. Second, our participants were self-selected to participate in the weight management programs, indicating their motivation and intention (and by association, perhaps their parents’ willingness to provide support) was higher than average. Thus, our results might be skewed toward highly motivated adolescents who are willing to use and co-use CHT with their parents and health care providers. Finally, our study does not include all possible members of a social support network (eg, peers and other family members) that may fill or contribute to the co-use roles we have outlined or introduce additional co-use possibilities. Each of these limitations is a call for future research to extend the scope of this study.

**Conclusions**

Social influences, as envisioned in our study (ie, the integration of the UTAUT focus on perceptions of others with the OCM’s consideration of patient and family engagement and empowerment) can affect the use of CHT as part of weight loss self-management. Our study investigated the desired CHT design features and co-user roles that can support not only the primary user but also the support environment (through parent co-use) and medical system support (through provider co-use) in self-management of adolescent overweight.

This study confirmed the wide applicability of the UTAUT model in a consumer health informatics context while recognizing the importance of co-use in the environmental and medical system components of the OCM. We found a strong willingness among adolescents to adopt co-use CHT features that assist with motivation, monitoring, and rewarding features in conjunction with parents’ and health care providers’ participation. Furthermore, we demonstrate that the expected roles and requirements of CHT co-use correspond to each other’s needs in the care model, whereas differences in the way CHT is co-used is transparent.

The study established differing features to encourage intention of use by each of the co-users but congruence in defining the primary and co-users’ roles. All co-users recognized the adolescent as the primary user with control of input and information in concert with concerns for privacy.
Overall, this study promotes awareness that the design of CHT can link multiple users to fulfill their roles in the management of adolescent overweight through complimentary and shared care tasks. The results of this study should prompt researchers and developers to recognize that CHT focused on adolescent overweight and obesity, and arguably many CHTs targeted to support the management of chronic conditions, should be designed, developed, and tested with the awareness of co-use to reinforce the social influences required for effective self-management.

Acknowledgments
This study was funded in part by the Saint Louis University’s Center for Sustainability grant. The authors greatly appreciate the work of Jean Hueelsing, Camp Jump Start (Living Well Foundation) founder, for her cooperation in the study. The authors also acknowledge Dr Amy Knoblock-Hahn (dietitian) and Dr Toree Malasanos (pediatric endocrinologist) for their collaboration and insight on aspects of the broader research project that seeded this study.

Conflicts of Interest
None declared.

References


Abbreviations

CCM: chronic care model
CHT: consumer health technology
mHealth: mobile health
OCM: obesity care model
UTAUT: unified theory of acceptance and use of technology

©Cynthia M LeRouge, Hyeyoung Hah, Gloria J Deckard, Haoqiang Jiang. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 29.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic
information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Feasibility and Acceptability of a Text Message–Based Intervention to Reduce Overuse of Alcohol in Emergency Department Patients: Controlled Proof-of-Concept Trial

Elizabeth Burner1,2, MD, MPH, MSc; Mark Zhang1,2, MD; Sophie Terp1,2,3, MD, MPH; Kelsey Ford Bench1,2, MD; Joshua Lee1,2, MD; Chun Nok Lam1,2, MPH, PhD; Jesus R Torres4, MD, MPH; Michael Menchine1,2,3, MD, MPH; Sanjay Arora1,2,3, MD
1Department of Emergency Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
2Los Angeles County + University of Southern California Medical Center, Los Angeles, CA, United States
3Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, CA, United States
4School of Medicine, University of California San Francisco, San Francisco, CA, United States

Corresponding Author:
Elizabeth Burner, MD, MPH, MSc
Department of Emergency Medicine
Keck School of Medicine
University of Southern California
1200 N State Street
Los Angeles, CA, 90033
United States
Phone: 1 323 409 6638
Email: eburner@usc.edu

Abstract

Background: Emergency department (ED) patients have high rates of risky alcohol use, and an ED visit offers an opportunity to intervene. ED-based screening, brief intervention, and referral to treatment (SBIRT) reduces alcohol use and health care costs. Mobile health (mHealth) interventions may expand the impact of SBIR Ts but are understudied in low-resource ED populations.

Objective: The objective of this study was to assess the feasibility of and patient satisfaction with a text-based mHealth extension of an ED screening program to reduce risky alcohol use in low-income, urban patients.

Methods: Research assistants screened a convenience sample of ED patients in person for risky alcohol use via the Alcohol Use Disorders Identification Test (AUDIT). Patients who reported AUDIT scores ≥8 and <20 were informed of their AUDIT score and risk. RAs invited patients with SMS text message–capable phones to receive mROAD (mobilizing to Reduce Overuse of Alcohol in the ED), an SMS text message–based extension of the ED screening program. mROAD is a 7-day program of twice-daily SMS text messages based on the National Institutes of Health’s Rethinking Drinking campaign. Participants were allocated to a control group (daily sham text messages without specific guidance on behaviors, such as “Thanks for taking part!”) or to the mROAD intervention group. Patients were interviewed at 30 days to assess acceptability, satisfaction, and changes in drinking behavior. Satisfaction was examined descriptively. Pre and post measurements of drinking behaviors and motivation were compared, as were differences in change scores between the intervention arms.

Results: Of 1028 patients screened, 95 (9.2%) exhibited risky alcohol use based on AUDIT, and 23/95 (24%) of those patients did not own an SMS text messaging–capable phone; this left 72/95 (76%) eligible patients. Among eligible participants, 48/72 (67%) agreed to enroll; 31/48 (65%) achieved follow-up (18/24 (75%) in the intervention group and 13/24 (55%) in the control group). Participants who completed follow-up reported high satisfaction. Changes in behavior were similar between the arms. Overall, the number of drinking days reported in the prior 30 days decreased by 5.0 (95% CI 1.7-8.3; P=.004), and the number of heavy drinking days decreased by 4.1 (95% CI 1.0 to 7.15, P=.01). Patients reported an 11-point increase (95% CI 2.6-20, P=.01, 10% overall increase) in motivation to change alcohol use via the Change Questionnaire. There were no statistical differences in drinking days, heavy drinking days, or motivation to change between the arms.

Conclusions: The mROAD trial was feasible. Over three-quarters of ED patients with risky alcohol use owned a text message–capable phone, and two-thirds of these patients were willing to participate; only 1 patient opted out of the intervention.
Although 35% of patients were lost to follow-up at 30 days, those patients who did follow up had favorable impressions of the program; more than 90% reported that SMS text messages were a “good way to teach,” and 89% of intervention arm participants enjoyed the program and found that the messages were motivating. Both the mROAD and sham message groups showed promising changes in alcohol use and motivation to change. mROAD is a feasible intervention that may reduce rates of risky alcohol use in ED patients.

**Trial Registration:** ClinicalTrials.gov NCT02158949; https://clinicaltrials.gov/ct2/show/NCT02158949

**KEYWORDS**
mhealth; alcohol misuse; emergency department; alcohol intervention

**Introduction**

Alcohol-related harm is responsible for an estimated $249 billion in yearly economic costs in the United States [1,2]. Efforts to reduce risky alcohol use are consistently cost-effective [3,4]. Unfortunately, only 10%–15% of patients who require treatment for substance use actually receive it [5,6]. Screening, brief intervention, and referral to treatment (SBIRT) is a cost-effective strategy to intervene in risky alcohol use. SBIRTs consist of screening (screening patients’ alcohol use with standardized tools), brief intervention for moderate risk screens (informing patients of the health risks posed by their current use and motivational interviewing techniques to encourage behavior changes), and referral to treatment for high-risk screens [7-9]. SBIRTs decrease ED utilization, health care costs, and risky alcohol use in multiple settings [10-14].

While SBIRTs can reduce risky alcohol use and decrease health care costs, many EDs still do not deploy them. Barriers of competing time priorities, lack of provider training in addiction medicine, and social stigma limit broad implementation of SBIRTs [15-18]. A brief intervention of 5-20 minutes per at-risk patient is recommended, which may exceed the total time a busy ED provider has to spend with a patient on that visit [19]. Strategies to reduce the provider time needed to deliver SBIRTs may increase their use and implementation.

Mobile health (mHealth) is a strategy that can be implemented to decrease the provider time required to deliver SBIRTs. mHealth includes the use of phone applications, SMS text messaging, and web-linked portals to provide public health interventions and clinical care. mHealth has been successfully used in the past to improve outcomes in a wide range of health issues, including substance use [20-22]. mHealth-based interventions for ED patients with risky alcohol use have been successful with young adult binge drinkers as well as with injured patients in New Zealand [23,24]. However, there is limited data on the role of mHealth in SBIRTs in other populations, including low-income, urban EDs and non–English-speaking patients. Previous work with low-income, urban ED patients has shown these patients to be ready to accept mHealth interventions and to own mobile phones capable of receiving these interventions [25-29]. However, it is not known if patients with high-risk alcohol use have the same access to mobile technology.

To understand the feasibility of an mHealth extension of an SBIRT in a low-income, predominantly non–English-speaking population, we conducted a proof-of-concept trial in the ED of an urban, academic safety net hospital. Patients received screening and notification of risk in the ED and were allocated to either an intervention group, which received twice daily theory-driven SMS text messages, or an active control group, which received daily nonspecific text messages for seven days. We collected feasibility data, perceptions of acceptability, and preliminary efficacy data 30 days after the intervention ended.

**Methods**

**Study Design and Location**

This quasiexperimental trial took place at an urban Level 1 trauma medical center (Los Angeles County + University of Southern California Medical Center) in Los Angeles, CA. Local institutional review board approval was obtained prior to the beginning of the study. The study was registered with ClinicalTrials.gov (NCT02158949).

**Screening and Recruitment**

Trained research assistants reviewed the ED electronic tracking board in real time between 7 AM and 11 PM over the course of 3 months (May through July 2014). Adult ED patients were approached in person for initial screening unless precluded by a clinical condition, language barrier (any language other than English or Spanish), or other inability to verbally consent to screening. Patients were screened using a tablet-based survey unless they preferred to have the survey read aloud to them by the research assistant. Patients were screened on level of alcohol use via the Alcohol Use Disorders Identification Test (AUDIT) developed by the World Health Organization (WHO) [30] as well as on mobile phone ownership and mobile technology use via questions developed by the Pew Center [25]. Patients with SMS text messaging–capable phones who were at risk for alcohol use disorders (AUDIT scores ≥ 8 but <20) were recruited to the study regardless of the reason for their visit [31]. AUDIT scores of 8-19 reflected patients who might benefit most from SBIRTs, while those with scores 20 and above required more intense intervention [32]. The AUDIT has excellent retest reliability, with a mean Cronbach α=.8 in a review of 10 studies [33]. Standard AUDIT scoring per WHO organization instructions was used [30].

At recruitment, RAs informed the patient that they were at risk for hazardous alcohol use and invited them to receive SMS text messages upon enrolling. Patients were offered an incentive of a $10 gift card if they choose to enroll to offset the cost of...
receiving the SMS text messages. Patients were given both a verbal explanation of the project and a copy of the informed consent form.

After agreeing to participate, participants were sequentially assigned to either the active control group or the intervention group. Next, patients were registered in the mHealth platform, which sent automated, unidirectional, broadcast SMS text messages for 1 week. Patients selected their preferred language for text messages as English or Spanish. Patients were not required to text a response to be enrolled.

**Measures**

On enrollment, patients reported their alcohol use by responding to two questions. The first question was a general alcohol use question: “Please think back over the last month. How many days did you drink?” The second question was a gender-based assessment of heavy drinking: “Over the last month, how many days did you drink heavily?” (“Heavy drinking” was defined for women as more than 3 drinks in one day and for men as more than 4 drinks in one day.) Patients were defined as everyday drinkers or binge drinkers based on their responses to the initial AUDIT screen. Additionally, participants reported their desire to change via the Change Questionnaire applied to risky alcohol use [34-36]. The Change Questionnaire consists of 12 statements on a patient’s belief of the importance of change, commitment to change, and ability to change; its Cronbach α=.86 when applied to alcohol use [37]. AUDIT scores and mobile phone ownership data were taken from the screening data.

**Intervention**

The mROAD intervention was entirely SMS text message–based, given the low access to advanced mobile technology in this population [27]. mROAD is a unidirectional, automated system; it was set to start the day after a patient enrolled in the trial. Messages were delivered on a Health Insurance Portability and Accountability Act (HIPAA)-compliant system that was compatible with the local pay-as-you-go cellular plans that are popular in this patient population. The mROAD intervention was developed in English and then translated to Spanish, with back translation by two native Spanish speakers to ensure a clear translation. Patients received either the mROAD intervention (described below) or a week of sham SMS text messages so that patient willingness to receive messages could be assessed in both arms.

The active control arm (sham) received a sham SMS text message greeting daily (e.g., “Thanks for taking part!”), while the intervention group (mROAD) received two text messages about alcohol use daily for 7 days. The intervention SMS text messages were adapted from the National Institutes of Health publication *Rethinking Drinking* [38]. The messages were shortened to fit the character limit of SMS text messages. The selected content included the consequences of drinking, motivational statements, and resources on how to obtain help to reduce drinking. Social norms theory and motivational interviewing strategies were emphasized, as supported by systematic reviews of the literature [39]. For example, social norms theory–based messages described normal drinking behavior, while motivational interviewing–based messages prompted participants to set a goal and write it down or type it out (see Figure 1 for example mROAD and sham messages).
Follow-up Procedures

Patients were contacted 30 days after the intervention to complete a telephone follow-up assessment. Research assistants were blinded to the treatment group at time of follow-up. Patients reported their days of drinking alcohol over the past month, days of heavy drinking over the past month, and desire to change their alcohol-drinking behavior by responding to the same questions asked at enrollment but without visual prompts. Patients in each group also completed a brief acceptability questionnaire. At the end of the trial, we collected the service records of the mobile health platform to ensure that the messages were delivered to each participant as scheduled.

Outcomes and Analysis

Feasibility was defined as >60% of eligible patients consenting and enrolling in the program and achieving 60% follow-up with participants. Previous studies with SBIRTs among ED patients yielded enrollment rates of 38%-87%, with follow-up rates between 49% and 89% [14,24,40]. For the few SMS text message–based alcohol interventions from the ED, follow-up rates have been between 75% and 82% [23,41,42]. However, previous work at our study site showed a maximum telephone follow-up rate of 70% for all comers [43]. To account for the lower anticipated follow-up rates for a low-income, non–English-speaking ED population in addition to patients with risky alcohol use, we determined 60% to be an acceptable follow-up rate.

We defined acceptability as greater than 90% of patients completing the 7 days of text messages without opting out by review of the mHealth platform service records. Our secondary acceptability outcome benchmark was 75% of participants agreeing with each statement in a brief, locally developed acceptability questionnaire. We also reported preliminary efficacy results with changes in days drinking alcohol, days heavily drinking, and desire to change drinking behavior. Preliminary efficacy results were compared to baseline and between groups with two-sample $t$ tests without assumption of equal variance.

Results

Feasibility Outcomes

2195 patients were identified by real time electronic tracking board review; 1167 could not be screened (see Figure 2 for
The patients enrolled in the study were predominately male, Latino, and aged 30-39 years (see Table 1 for the complete study population characteristics). More than half spoke Spanish primarily. Compared to patients in the mROAD arm, patients allocated to the sham message intervention group had lower self-reported rates of mental illness (13% vs 25%) and higher numbers of days drinking alcohol and days drinking heavily in the prior month. These baseline differences were not statistically significant.

Nearly two-thirds (31/48, 65%) of enrolled patients were successfully reached for follow-up; follow-up was higher in the intervention group (18/24, 75%) than in the control group (13/24, 54%). More patients in the intervention group reported receiving messages (17/18, 94%) than patients in the active control group (11/13, 85%).
Table 1. Baseline characteristics of the study participants (N=48).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sham group (n=24)</th>
<th>mROAD\textsuperscript{a} group (n=24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>35.4 (10.6)</td>
<td>38.8 (13.5)</td>
</tr>
<tr>
<td>Male gender, n (%)</td>
<td>21 (88)</td>
<td>19 (79)</td>
</tr>
<tr>
<td>Spanish speaking, n (%)</td>
<td>14 (58)</td>
<td>14 (58)</td>
</tr>
<tr>
<td>Latino ethnicity, n (%)</td>
<td>21 (88)</td>
<td>17 (71)</td>
</tr>
<tr>
<td>Self-report of mental illness, n (%)</td>
<td>3 (13)</td>
<td>6 (25)</td>
</tr>
<tr>
<td>Number of days drinking alcohol last month, mean (SD)</td>
<td>11.5 (8.4)</td>
<td>7.6 (8.6)</td>
</tr>
<tr>
<td>Number of days drinking heavily\textsuperscript{b} last month, mean (SD)</td>
<td>5.5 (7.13)</td>
<td>4.9 (7.2)</td>
</tr>
<tr>
<td>Motivation to change drinking (score 0-120), mean (SD)\textsuperscript{c}</td>
<td>89.3 (33.8)</td>
<td>88.8 (6.7)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}mROAD: mobilizing to Reduce Overuse of Alcohol in the emergency Department.

\textsuperscript{b}Drinking heavily was defined as >3 standard-sized drinks per episode for women and >4 standard-sized drinks per episode for men.

\textsuperscript{c}Measured by the Change Questionnaire.

**Acceptability Outcomes**

Overall, acceptance of the intervention was high among patients who were followed up, and all acceptability benchmarks were achieved. Review of the mHealth platform records indicated that only 1 patient in the intervention arm opted out of the messages; the patient opted out after the first intervention message was sent. All other patients received the full 7 days of messages. Of the 31 patients assessed at follow-up, more than 90% agreed that using SMS text messages was a “good way to teach,” nearly four-fifths reported that the number of messages per day was “just right,” and more than half wanted the messages to continue (see Table 2).

There were differences in acceptability between the two arms. Patients in the intervention arm had higher acceptance of the program than patients in the sham arm. More patients in the intervention arm enjoyed the program, were willing to recommend it to friends and family, were motivated by the messages, and perceived that the messages came at the right time of day.

**Preliminary Efficacy Outcomes**

Patients in both arms reported increased motivation to change drinking behavior, decreased days drinking any alcohol, and decreased days drinking heavily (see Table 3). Overall, participants reported increased motivation to change alcohol use, with an 11-point increase (95% CI 2.6-20, \( P = .01\), 10% overall increase) on the Change Questionnaire. The number of reported drinking days in the prior 30 days decreased by 5 (95% CI 1.7-8.3, \( P = .004\)) and heavy drinking days decreased by 4.1 (95% CI 1.0-7.15, \( P = .01\)). The differences in the changes between the arms were not significant; however, the sham message arm overall trended toward larger improvements.

Table 2. Acceptability as indicated by participants’ agreement with the following statements at 30 day follow-up, n (%).

<table>
<thead>
<tr>
<th>Statement</th>
<th>Sham group (n=13)</th>
<th>mROAD\textsuperscript{a} group (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using texts is a good way to teach</td>
<td>12 (92)</td>
<td>17 (94)</td>
</tr>
<tr>
<td>I enjoyed the mROAD program</td>
<td>9 (69)</td>
<td>16 (89)</td>
</tr>
<tr>
<td>I would like the messages to continue</td>
<td>7 (54)</td>
<td>10 (56)</td>
</tr>
<tr>
<td>I would recommend mROAD to family and friends</td>
<td>9 (69)</td>
<td>16 (89)</td>
</tr>
<tr>
<td>I was motivated by the mROAD messages</td>
<td>9 (69)</td>
<td>16 (89)</td>
</tr>
<tr>
<td>The messages came at a good time for me</td>
<td>9 (69)</td>
<td>15 (83)</td>
</tr>
<tr>
<td>The number of messages per day was just right</td>
<td>10 (77)</td>
<td>14 (78)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}mROAD: mobilizing to Reduce Overuse of Alcohol in the emergency Department.
they were more likely to recommend it to a friend or family, higher satisfaction with and motivation from the program, and received theory-based mROAD messages reported generally of in-ED screening to be helpful and motivating. Patients who engaged with the study; 75% were reached for follow-up,

with two-thirds of patients following up by phone; this indicates 

that screening and enrollment was feasible; more than 60% of patients with risky alcohol intake owned an SMS text message–capable phone, and more than 60% of eligible patients agreed to participate. In addition, patients accepted the mROAD intervention; only 1 patient left the intervention, and more than 80% of intervention group patients who were followed up reported favorable perceptions of mROAD. Preliminary efficacy results from the combined groups indicate that mHealth extensions of SBIRTs should be tested in larger and longer trials.

As a proof-of-concept trial, this study shows that mHealth extensions of SBIRTs are feasible in a low-income, urban, predominantly Latino ED patients. We found that screening and enrollment was feasible; more than 60% of patients with risky alcohol intake owned an SMS text message–capable phone, and more than 60% of eligible patients agreed to participate. In addition, patients accepted the mROAD intervention; only 1 patient left the intervention, and more than 80% of intervention group patients who were followed up reported favorable perceptions of mROAD. Preliminary efficacy results from the combined groups indicate that mHealth extensions of SBIRTs should be tested in larger and longer trials.

Table 3. Mean changes in drinking habits and desire to change for each group and combined among participants who were followed up.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sham group (n=13)</th>
<th>mROAD group (n=18)</th>
<th>Combined (n=31)</th>
<th>Differences between groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95% CI)</td>
<td>P value</td>
<td>Mean (95% CI)</td>
<td>P value</td>
</tr>
<tr>
<td>Decrease in days drinking alcohol</td>
<td>8.5 (3.0 to 13.9)</td>
<td>.005</td>
<td>2.5 (–1.7 to 6.7)</td>
<td>.23</td>
</tr>
<tr>
<td>Decrease in days drinking heavily</td>
<td>5.6 (–0.1 to 11.4)</td>
<td>.05</td>
<td>3 (–0.7 to 6.7)</td>
<td>.11</td>
</tr>
<tr>
<td>Increased motivation to change (score 0-120)</td>
<td>16.6 (–1.6 to 34)</td>
<td>.07</td>
<td>7.6 (–1.7 to 17)</td>
<td>.01</td>
</tr>
</tbody>
</table>

a mROAD: mobilizing to Reduce Overuse of Alcohol in the emergency Department.
b tests were used to compare pre-post measures and between-group differences.
c Heavy drinking was defined as >3 standard-sized drinks per episode for women and >4 standard-sized drinks per episode for men.
d Measured by the Change Questionnaire.

Discussion

We conducted this proof-of-concept trial to assess the feasibility of an mHealth extension of an ED-based SBIRT to decrease risky alcohol intake among low-income, urban, predominantly Latino ED patients. We found that screening and enrollment was feasible; more than 60% of patients with risky alcohol intake owned an SMS text message–capable phone, and more than 60% of eligible patients agreed to participate. In addition, patients accepted the mROAD intervention; only 1 patient left the intervention, and more than 80% of intervention group patients who were followed up reported favorable perceptions of mROAD. Preliminary efficacy results from the combined groups indicate that mHealth extensions of SBIRTs should be tested in larger and longer trials.

As a proof-of-concept trial, this study shows that mHealth extensions of SBIRTs are feasible in a low-income, urban, predominantly Latino population. As with most alcohol screening programs, the most common reason for ineligibility was not reporting risky levels of alcohol intake. Approximately 30% of patients with risky alcohol use were ineligible due to lack of mobile phone ownership; this rate is higher than in similar interventions with younger, nonminority populations [44]. As the mobile capacity of low-income Latino patients continues to increase, the gap between patient capacity and intervention delivery will narrow [26,45]. Additionally, more than two-thirds of eligible patients chose to participate, which is at the high end of mHealth interventions for alcohol use [24,40,44]. Lastly, we maintained an adequate follow-up rate, with two-thirds of patients following up by phone; this indicates that studying this type of intervention is feasible.

In addition to being feasible, this mHealth extension of in-ED screening was acceptable and satisfactory to patients. Patients who received theory-based mROAD messages were more engaged with the study; 75% were reached for follow-up, compared with 54% of patients who received daily sham messages. Patients in both groups found the mHealth extension of in-ED screening to be helpful and motivating. Patients who received theory-based mROAD messages reported generally higher satisfaction with and motivation from the program, and they were more likely to recommend it to a friend or family member. These acceptability results are promising for larger-scale trials and widespread implementation.

While mHealth extensions of in-ED SBIRTs are feasible, most EDs still do not conduct standardized screening and intervention, which limits their implementation [18]. However, promising work using patient self-administered and computer-based screening and notification of risk provides an opportunity to increase the number of patients screened and referred to an mHealth extension of an SBIRT [46-49]. As more EDs move to self-administered screening of behavioral risk factors and social determinants of health via computer, tablet, and mobile device interfaces, it may be possible to formally screen more patients for risky alcohol use [50,51]. Integrating formalized screening for alcohol behaviors increases implementation of screening and SBIRTs [52,53]. Increased screening could provide a larger target population, which could require increased resources at individual institutions. By using mHealth SBIRT strategies in combination with in-ED computer-based, tablet-based, and mobile device–based screening, the scope of SBIRTs can be increased. For clinics and EDs that already screen for risky alcohol use, similar mHealth extensions of screenings and brief interventions would require marginal extra workforce time. mHealth interventions hold potential to create large-scale programs to reduce risky drinking among ED patients without increasing demands on an already overstretched ED workforce.

In our study, patients in the sham message arm reported decreased drinking days, heavy drinking days and increased motivation to change drinking. The patients in the sham message arm started with higher reported drinking days, which correlates with larger decreases in risky alcohol behavior in prior ED-based SBIRTs [54,55]. Additionally, the follow-up rate was lower in the sham message arm; sham message patients who were followed up may be more motivated than the average ED patient. While the difference was not significantly different from the theory-based message arm, there are several possible explanations if this finding is verified in fully powered studies. Patients receiving theory-based messages may become more aware of their drinking habits if they are reminded with messages pertaining to their drinking rather than sham messages alone. As a consequence, they may more accurately report their drinking frequency than the patients in the control group. This
study did not have a usual care control group, as all patients were first informed in the ED during their initial contact that they were at risk for alcohol abuse. The minor intervention of daily SMS text messages linked to the ED-based screening may have promoted a change in the patients’ habits.

While this proof-of-concept study is promising, it has several limitations. A strength of this study and of the mROAD program is the demonstration of the potential of a simple, easily scalable, automated system to encourage positive behavior changes; however, the small sample size, quasiexperimental design, and short follow-up period prevent conclusions about sustained behavior changes or differences between patients who received theory-driven vs sham messages. Patients in the sham message and theory-driven mROAD arms had similar reported changes in alcohol use and motivation to change; this indicates that either the sham messages after the in-ED screening and risk notification had beneficial effects alone or that the natural history of an ED visit may include a decrease in alcohol use. Further study of this type of intervention may require a control group with less activation. This study was conducted at a single site, which may limit the generalizability of the feasibility findings. Additionally, the logistics of patient follow-up from an ED-based study that serves a low-income, non–English-speaking population creates potential for biased results due to differential follow-up.

This proof-of-concept study shows that low-income, urban ED patients can feasibly be enrolled in mHealth extensions of ED-based screening and brief intervention programs. We found that patients were willing to participate in mROAD and were accessible for follow-up. mHealth extensions of face-to-face clinical care can extend the impact of an ED visit well beyond the physical confines of the hospital.

**Acknowledgments**

This project was supported by a USC Department of Emergency Medicine Seed Grant, a Keck School of USC Medicine Medical Student Summer Research Award, and a Keck School of Medicine of USC Dean’s Scholar Award. ST’s time was supported by an Individual Postdoctoral Fellowship Award (F32 HS022402-01) from the Agency for Healthcare Research and Quality during the study period. We wish to thank Carla Martinez for her assistance in editing and table preparation.

**Authors’ Contributions**

EB designed the intervention and study, led the data analysis, and takes full responsibility for the accuracy of the work. MZ contributed to the data analysis, wrote the first draft of the manuscript, and participated in the revision of the final manuscript. MZ, KFB, JL, and JRT led the data collection, designed and implemented the follow-up strategies, and participated in drafting and editing the final manuscript. CNL contributed to and conducted the data analysis plan. MM, ST, and SA contributed equally as senior mentors to the design of the study and to drafting and editing the manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 321 KB - mhealth_v8i6e17557_app1.pdf ]

**References**


41. Suffoletto B, Chung T, Muench F, Monti P, Clark DB. A Text Message Intervention with Adaptive Goal Support to Reduce Alcohol Consumption Among Non-Treatment-Seeking Young Adults: Non-Randomized Clinical Trial with Voluntary Length of Enrollment. JMIR Mhealth Uhealth 2018 Feb 16;6(2):e35 [FREE Full text] [doi: 10.2196/mhealth.8530] [Medline: 29453191]


Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
ED: emergency department
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
mROAD: mobilizing to Reduce Overuse of Alcohol in the emergency Department
SBIRT: Screening, Brief Intervention, and Referral to Treatment
WHO: World Health Organization

©Elizabeth Burner, Mark Zhang, Sophie Terp, Kelsey Ford Bench, Joshua Lee, Chun Nok Lam, Jesus R Torres, Michael Menchine, Sanjay Arora. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 04.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Developing a Text Messaging Intervention to Reduce Deliberate Self-Harm in Chinese Adolescents: Qualitative Study

Suqian Duan1,2, MRes; Haoran Wang1, MEd; Amanda Wilson3, PhD; Jiexi Qiu1, BSc; Guanmei Chen1, BSc; Yuqiong He1, MBBS; Yuanyuan Wang3, PhD; Jianjun Ou1*, MD, PhD; Runsen Chen1,4*, MBMS

1Department of Psychiatry, The Second Xiangya Hospital, Central South University, Changsha, China
2Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, United Kingdom
3Division of Psychology, Faculty of Health and Life Sciences, De Montfort University, Leicester, United Kingdom
4Department of Psychiatry, University of Oxford, Oxford, United Kingdom
*these authors contributed equally

Corresponding Author:
Runsen Chen, MBMS
Department of Psychiatry
The Second Xiangya Hospital
Central South University
139 Renmin Middle Rd, Yuan Jia Ling Shang Quan, Furong District
Changsha
China
Phone: 86 731 85292158
Email: Runsen.chen@psych.ox.ac.uk

Abstract

Background: Deliberate self-harm is common during adolescence and can have detrimental consequences for the well-being of adolescents. Although it is sometimes difficult to engage adolescents in traditional psychotherapies for deliberate self-harm, SMS text messaging has been shown to be promising for cost-effective and low-intensity interventions.

Objective: This study aimed to investigate the views of Chinese adolescents with deliberate self-harm about SMS text messaging interventions in order to develop an acceptable and culturally competent intervention for adolescents with deliberate self-harm.

Methods: Semistructured interviews were conducted with 23 adolescents who had experience with deliberate self-harm. The transcripts of the interviews were analyzed using thematic analysis.

Results: Four themes were identified: beneficial perception of receiving messages, short frequency and duration of messages, caring content in messages, and specific times for sending messages. Most of the participants perceived SMS text messaging interventions to be beneficial. The key factors that emerged for the content of the intervention included encouragement and company, feeling like a virtual friend, providing coping strategies, and individualized messages. In addition, the preferred frequency and duration of the SMS text messaging intervention were identified.

Conclusions: Our study will help in the development of a culturally appropriate SMS text messaging intervention for adolescents with deliberate self-harm. It has the potential to decrease deliberate self-harm instances by providing acceptable support for adolescents with deliberate self-harm who may be reluctant to seek face-to-face psychotherapies.

(JMIR Mhealth Uhealth 2020;8(6):e16963) doi:10.2196/16963

KEYWORDS
text messaging; deliberate self-harm; adolescents; qualitative research; mental health

Introduction

Deliberate self-harm is defined as any nonfatal act of self-poisoning or self-injuries irrespective of the extent of suicidal intent or any other type of motivation [1]. The prevalence rate of deliberate self-harm is very high throughout adolescence, with percentages fluctuating between 12% and 23% [2]. Deliberate self-harm has been associated with a wide range of physical and psychiatric conditions, such as depression and suicidal behaviors, resulting in detrimental outcomes for the well-being of adolescents with deliberate self-harm [3]. Despite this, only a few interventions have been designed...
specifically for adolescents with deliberate self-harm, and of them, few have demonstrated promising therapeutic outcomes [2]. Thus, there is an urgent need to develop novel and acceptable interventions that target deliberate self-harm among adolescents.

One of the challenges in developing interventions for adolescents with deliberate self-harm is the reluctance of adolescents to engage with health professionals, especially in countries where mental health services are underdeveloped. Compared with traditional psychotherapies such as cognitive behavioral therapy, cost-effective and low-intensity treatment might be more feasible and acceptable among adolescents [4].

Recently, researchers have developed lower cost, higher coverage interventions, such as SMS text messaging interventions [5]. Texting is highly valued among adolescents nowadays, not only as a means of socializing but also as a way of providing and gaining peer support when facing difficulties in life [6]. Thus, SMS text messaging interventions might overcome some barriers of traditional psychotherapies and engage adolescents who are unwilling to attend face-to-face treatment [4].

Prior work has demonstrated the potential of SMS text messaging interventions as a cost-effective psychotherapy in the field of mood disorders, self-harm, and suicide-related behaviors with somewhat inconsistent findings [5,7]. For example, a randomized clinical trial suggested that participants who received SMS text messaging in the form of caring contact for 12 months had lower odds of experiencing any suicidal ideation than those who received standard care alone [5]. Furthermore, Hull et al [8] found that text interventions were effective in reducing symptoms of depression and anxiety to the extent of a diagnostic change. On the other hand, SMS text messaging interventions have had a significant effect on the likelihood or severity of current suicidal ideation [5] but has been found ineffective in improving depressive symptoms in another study [7]. Although the reasons behind these controversies remain unclear, all the studies have used very different content, and the frequency and duration of the text messages differed, which could partially explain the inconsistent findings. More importantly, none of these studies have developed interventions based on the perceptions and preferences of service users. According to Ram et al [9], service user engagement increases the possibility of developing interventions that are considered safe, have better usability, show they are clinically effective, and are culturally competent.

An important way of involving service users in the design of an intervention is to adopt a qualitative method, such as SMS text messaging interventions [4,6]. Given the importance of involving service users in the design of an intervention, this study aimed to investigate the lived experiences of Chinese adolescents with deliberate self-harm regarding SMS text messaging intervention in order to develop an acceptable and culturally tailored intervention. Semistructured interviews were conducted to gather information from potential service users of an SMS text messaging intervention.

**Methods**

**Participants**

This study recruited participants from both the outpatient and inpatient units of the Second Xiangya Hospital, with special consideration given that inpatients were stable enough to participate. The inclusion criteria were as follows: (1) they must be between 12 and 17 years old, (2) they must have normal intelligence, (3) they should be able to complete the interview, (4) they have self-injured twice or more in the past 6 months, (5) they have no severe neurological or organic diseases, and (6) their parent or guardian understands the information sheet and agrees to fill in a consent form.

**Procedure**

Participants were invited to participate in one semistructured in-depth interview with the researcher alone. At the beginning of the interviews, participants’ diagnoses of psychiatric disorders were confirmed with the clinical staff. Their demographic information and methods of self-harm were both self-reported and confirmed with the clinical staff. The interview started with an introduction of SMS text messaging interventions and some examples of text messages. These examples were designed based on SMS text messaging used in previous studies [6,10]. In addition, the examples were also discussed and reviewed by senior experts (clinical psychologists and psychiatrists) to tailor the texts for Chinese adolescents with deliberate self-harm. Example messages were only chosen if they were judged to be able to reduce the impulses of deliberate self-harm or negative emotions in a Chinese context. All example messages are listed in Textbox 1.

The interview was guided by open-ended questions and designed to understand the perceptions of participants on SMS text messaging interventions. The aim was to gain their perspectives on how to reduce their deliberate self-harm. An example of an interview question is While you are having treatment for your self-harming behavior, we want to send you some text messages to help reduce your urges to self-harm. These messages could include information regarding how to deal with your low mood. How would receiving these messages make you feel? The interview was developed based on questions identified during the literature review as well as previous text-base on interventions that have been shown to be effective [6]. All interviews were conducted in the psychotherapy room in the hospital by 3 authors (SD, JQ, and HW) with a clinical psychology background, who were trained by a senior clinical psychologist (RC). All interview questions and answers were communicated in Mandarin Chinese and then translated into English by the first author. Each interview lasted approximately 30 min; this was done to ensure that the participants did not become distressed. At the end of the interview, participants...
were asked to choose the messages they liked the most and the least from the sample provided.

**Textbox 1.** Examples of messages shown to participants at the interview.

<table>
<thead>
<tr>
<th>Number</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nice to meet you! In the coming months, I will be honored to share your joy and pain. From today, please treat yourself as if you were someone else, and do some sweet little things for yourself, such as: light candles in the room, and find the smell that makes you happy. If you want to get more help for self-injury contact us. Have a good day!</td>
</tr>
<tr>
<td>2</td>
<td>What lies behind you and what lies before you are tiny matters compared to what lies within you. Have faith in yourself and success can be yours. Hope you have a lovely week.</td>
</tr>
<tr>
<td>3</td>
<td>Hello. When you are in a bad mood, you may think: “I have experienced many difficult situations and I survived, this time is no exception.” Hope you have a good day!</td>
</tr>
<tr>
<td>4</td>
<td>Letting go of resentment is a gift you give yourself, and it will ease your journey immeasurably. Make peace with everyone and happiness will be yours. Hope everything is going well with you.</td>
</tr>
<tr>
<td>5</td>
<td>If you want to self-harm, you may try the following things instead: (1) Draw the person or thing you hate on the balloon and then pop it. (2) Write a letter to someone you hate or who hurts you, talk about what they did to you and why you hate them. Save the letter and read it later. I hope your life is getting better and better</td>
</tr>
<tr>
<td>6</td>
<td>Pay attention to activities that have a positive impact on your mood. Note these activities and refer to them when you hit a low point to improve your mood. Remember to take good care of yourself!</td>
</tr>
<tr>
<td>7</td>
<td>Hi! Remember, doing something pleasant is the best way to get rid of pain. You may try the following things: (1) Eat chocolate or your favorite food. (2) Do something exciting, such as learning a sport. (3) Take a movie or video with your camera. Hope you have a good day!</td>
</tr>
<tr>
<td>8</td>
<td>Hi, do you remember me? If you are in a low mood, you may be able to call the self-injury crisis hotline and talk to people. The number is: 4001619995. I hope everything is OK recently! Don’t worry about yesterday and tomorrow, we only live for today! Wish you a happy life!</td>
</tr>
<tr>
<td>9</td>
<td>Today is your birthday. Happy birthday to you! In order to reward yourself for overcoming difficulties and treating yourself well in the past year, you may go to your favorite coffee shop for tea or coffee with your friends.</td>
</tr>
</tbody>
</table>

**Data Analysis**

All interviews were audio recorded and then transcribed verbatim by 2 authors (SD and JQ). The data were then analyzed using thematic analysis. According to Braun and Clarke [11], there are 5 phases included in this approach. First, 2 authors familiarize themselves with the transcripts by independently reading and rereading them. Second, the initial codes of the transcripts are generated. Third, themes are identified from the transcripts. Fourth, the 2 authors share their themes and compare them with each other. Fifth, a final consensus on the themes is achieved between the 2 authors to decide how to define and name themes [12,13]. After this, the 2 authors engaged in a discussion with a senior investigator (RC), and the senior investigator read over all the themes and subthemes to assess the reliability of the findings.

**Ethics Approval**

This study protocol was approved by the Ethics Committee of the Second Xiangya Hospital of Central South University.
### Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Diagnosis</th>
<th>Methods of self-harm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>Male</td>
<td>Bipolar disorder</td>
<td>Cutting</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and scratching</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and head banging</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>Female</td>
<td>Depression and anxiety</td>
<td>Punching the wall</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and asphyxiation</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and burning</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>Female</td>
<td>Bipolar disorder</td>
<td>Cutting and head banging</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and dousing head with cold water</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>Female</td>
<td>Depression with psychotic symptoms</td>
<td>Cutting and head banging</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>15</td>
<td>17</td>
<td>Male</td>
<td>Depression</td>
<td>Cutting</td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>Male</td>
<td>Depression</td>
<td>Cutting and poisoning</td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting, biting, and pinching</td>
</tr>
<tr>
<td>18</td>
<td>15</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and pinching</td>
</tr>
<tr>
<td>19</td>
<td>17</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and head banging</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting, biting, and strangling</td>
</tr>
<tr>
<td>21</td>
<td>15</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting, pinching, and hair pulling</td>
</tr>
<tr>
<td>22</td>
<td>14</td>
<td>Female</td>
<td>Depression</td>
<td>Cutting and pinching</td>
</tr>
<tr>
<td>23</td>
<td>15</td>
<td>Male</td>
<td>Depression</td>
<td>Cutting and pinching</td>
</tr>
</tbody>
</table>

### Choice of Sample Messages

Several participants chose messages 7 and 9 as the messages they liked the most. Messages 5 and 8 were also chosen by some participants, whereas only a few participants chose messages 3, 4, and 6. In addition, 4 participants did not choose any message as their favorite. Message 4 was the least liked message among the majority of participants. Several participants did not like messages 2 and 3 as well. The number of participants and their choice of each message are summarized in Table 2.

### Table 2. Number of participants and their choice of messages.

<table>
<thead>
<tr>
<th>Message no.</th>
<th>Content of the message</th>
<th>Participants who liked the message, n</th>
<th>Participants who disliked the message, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treat yourself well</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Have faith in yourself</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>You will survive</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>Let go of resentment</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Try different things to DSHa</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Pay attention to positive things</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>Suggestions for activities</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>Self-harm hotline</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Birthday greeting</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>N/Ab</td>
<td>Did not choose any message</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

aDSH: deliberate self-harm.
bN/A: not applicable.
Thematic Analysis

The following sections represent the perceptions of the adolescents on SMS text messaging interventions in further detail. Four themes were identified from the transcripts of the semistructured interviews: (1) beneficial perception of receiving messages, (2) short frequency and duration of the messages, (3) caring content in messages, and (4) specific times for sending messages. Each section starts with the name of the theme. The extracts that best represent the theme are then presented and discussed.

Theme 1: Beneficial Perception of Receiving Messages

Most participants mentioned that the SMS text messaging intervention could be beneficial for them and were willing to receive text messages. Several of them said that the messages could help them have prompt self-reflection and stay calm:

...I hope to receive this type of messages because it helps me think about myself and my recent behaviours. It may ask me to take good care of myself, so I think I should treat myself well... [p13]

Other participants said that it could be helpful to know what resources are available, such as research-based information or a contact number of someone whom they could talk to when they want to self-harm, which might reduce their urges to self-harm:

...Having some scientific understanding of my symptoms via text-messages could be effective for reducing my self-harm. It is also helpful to have some useful information such as contact number of people who are there to help... [p16]

However, some participants also said that they would not like a contact number because they would not know the reliability of this information. In addition, there were also a few participants who doubted the benefit of the intervention, as they believed the messages would not improve their depressed mood to a large extent:

...I am not sure whether text-messaging intervention will be helpful or not. Reading these messages might be boring for me, but I believe it does no harm... [p22]

...When I read the messages, I will not have any particular feeling towards it, as I have read too much information similar to this. In addition, if there is information regarding certain website or certain number, I will think they are trying to advertise or blackmail me... [p19]

Furthermore, a small number of participants believed that they would not find this intervention useful because their parents keep their mobile phones for them most of the day. However, they mentioned that the intervention might also be helpful if their parents receive these messages.

Theme 2: Short Frequency and Duration of Messages

Participants had different suggestions for the frequency of text messages. Most participants agreed that it would be great to receive the messages more often than once a week but less often than once a day. This frequency would allow them to pay sufficient attention to the content of the messages:

...I prefer to receive this type of messages every three days. If I receive them too often, I will not want to read them... [p9]

...I am not sure about the frequency, but I would say do not be too frequent. Receiving messages once every two to three days might be good... [p22]

A few participants mentioned that they would like to receive messages more often than once a day, with 1 participant specifically mentioning 1 message each during the day and night. Importantly, 2 participants mentioned that text messages could be more frequent at the beginning, and then the frequency should reduce with time:

...Just keep sending the messages at the beginning of this intervention (with a high frequency). The frequency can decrease after some time... [p1]

In addition, some participants answered that they did not care about the frequency of the messages very much because the messages would be equally effective regardless of how often they receive the messages.

Regarding the duration of the intervention, many participants said it would be great to receive these messages for a duration of 2 months to 6 months. As they suggested, if the duration is too long, they would no longer pay attention to the contents. Moreover, the effectiveness might be reduced if the duration is too short:

...I think three-month time will be great for me, if I receive these for a longer duration, I will not read them... [p9]

...I suggest to send messages for one to two months, but it also depends on how long the patient recover from the symptoms... [p19]

Nevertheless, 2 participants suggested a long duration (1 year) and a short duration (1 month). Furthermore, 2 participants said that the duration could be as long as possible, as they would not want to stop receiving the messages:

...I would like to receive these messages for a duration as long as possible, if I stop receiving these messages one day, I will feel sad and empty... [p1]

Theme 3: Caring Content in Messages

There are 4 key factors identified in the caring content of messages: encouragement and company, feeling like a virtual friend, providing coping strategies, and individualized messages.

Many participants mentioned that they would like to receive messages as a form of social support. This is important, as it can make them feel encouraged and less isolated, which could potentially improve their depressive symptoms and low mood:

...The messages I want to receive may include any words for encouragement, help and comfort. These will make me feel supported and have the strength to fight with my disease... [p2]
...I need some encouragement and sense of being cared for in my life, so if the messages could include this type of information, I might feel better... [p5]
I cannot think about any particular type of messages, but I would say ‘I will always be with you’ will be helpful for me, it makes me feel less lonely... [p7]
...I guess something like “you are not alone” in the messages will make me feel better... [p11]

Although the participants used a tentative language, it was clear, regardless of whether they received a text message or not, that feeling lonely was at the core of deliberate self-harm. They wanted stable, long-term caring relationships to resolve this.

Some participants said that the messages should make them feel like they were sent by their friends, rather than by psychologists, parents, teachers, or an authority. Imagining that they were receiving messages from friends would make them feel less lonely. On the other hand, if the messages are too official, they might feel bored and might not want to look at the messages:

...Don’t include messages that make me feel like my parents are teaching me something, I need encouragement from a friend, not from a teacher or parent... [p10]
...The messages should not be very official. Instead, they need to be in the simple languages we use in daily lives. It is even better if it can make me feel like talking with a friend.... [p23]

A few participants suggested that the messages should include coping strategies and knowledge of self-harm. This is important for them to understand why they start self-harming and what they can do to prevent their self-harm:

...It is good for the messages to include some coping strategies for self-harm in my daily life. It will be very helpful if I receive this type of thing before I want to harm myself... [p19]
...The messages should have a list of coping strategies in simple languages. I will read them if I am interested... [p22]
...If they (the messages) let me know how I could deal with my stress; this might be useful... [p2]

Participants mentioned that the intervention could be more effective if the messages were tailored to give them a further sense of coming from someone caring. Participants specifically mentioned that generic messages that are sent to everybody would not work for them:

...I will enjoy reading these messages if I know they are tailored for me, rather than generic messages sent to everybody who has similar conditions... [p23]
...I will be very interested in the content of the messages if they are sent ‘one to one’, otherwise, I guess it does not have any personal meaning for me... [p8]

**Theme 4: Specific Times for Sending Messages**

As mentioned by the participants, receiving the messages can reduce their urges to self-harm. Most participants stated that their urges to self-harm were the strongest in the evening, especially when they were alone. However, a few participants also said that they harmed themselves in the afternoon or morning. A small number of participants stated that there was no fixed time for conducting self-harm; they would harm themselves whenever they felt depressed or irritable:

...I normally want to harm myself in the evening when I am alone, I just want to hurt myself because I feel sad... [p2]
...There was no fixed time for me to hurt myself, I will do it whenever my mood bursts... [p3]

Regarding how self-harm related to their life events, most participants said that they would self-harm when stressful life events occurred. These included being bullied, divorce of parents, difficult peer relationships, break-up in a romantic relationship, feeling unsupported, and feeling stressed from schoolwork. In addition, several participants identified that their urges to self-harm would increase before they return to school after the summer holidays. Nevertheless, 3 participants could not think of any particular events that triggered their self-harm:

I hurt myself when bad things happened in my life, such as something that makes me feel I am not loved by my parents and when I fail to have good performance in exams... [p1]
...I start to self-harm mainly after some bad things happened, such as being told off by my parents, being bullied in school etc... [p2]
...I will cut myself when I feel stressed at school and after I argue with my parents... [p23]
I harm myself every time I break up with my girlfriends, the severity of my self-harm depends on how much I care about her... [p4]

**Discussion**

**Principal Findings**

To our knowledge, this is the first study to investigate the views of adolescents with deliberate self-harm on SMS text messaging interventions to develop an acceptable and culturally tailored intervention. There were 23 patients interviewed at the psychiatric department who met the inclusion criteria. The analysis showed that by using a qualitative approach, most of the participants perceived the intervention as beneficial for them. In addition, the preferred frequency and duration of an SMS text messaging intervention were identified. It was found that the key time for delivering the messages is during the evening and after stressful events or during a crisis. Four key factors emerged for the content of the intervention: encouragement and company, feeling like a virtual friend, providing coping strategies, and individualized messages. On the basis of these results, SMS text messaging interventions tailored for Chinese adolescents will be developed, and randomized controlled trials assessing the feasibility and effectiveness of this intervention will be planned by our research team.
Perception of Receiving Messages

Despite different cultural contexts, the perception of SMS text messaging interventions is somewhat similar within the Western and Chinese cultures. Consistent with Owen et al [6], we found a broad range of support among service users for using text messages as an intervention for deliberate self-harm. Nevertheless, compared with Western studies, some participants in our study were hesitant toward receiving the SMS text messaging intervention, as they considered the effectiveness of this intervention unknown or questionable. Although the clear reason behind this is unknown, it could be related to the generally low engagement with and low level of understanding of mental health services among Chinese adolescents.

Frequency and Duration of Messages

Few previous studies have set out to understand what the appropriate frequency and duration are for SMS text messaging interventions from a service user’s perspective. Nevertheless, in a study by Larsen et al [4], participants indicated that frequent, but not too frequent, messages would be useful. This is consistent with our findings that most participants suggested a frequency of more than once a week but less often than once a day. In addition, our participants proposed a duration of intervention that is short rather than long, with most of them suggesting 2 to 6 months. This duration has been employed in most SMS text messaging interventions in the literature [14], and our findings provide supporting evidence for this design, confirming its acceptability from a service user’s perspective.

Content in Messages

Considering the content of a text messaging intervention is the most important step when designing the intervention. Compared with research conducted within the Western culture, our findings have demonstrated both consistency and novelty. To illustrate, our analysis identified some similar factors as those identified by 2 studies [4,15], such as encouragement and coping strategies to deal with deliberate self-harm. In addition, the factor individualized messages has also been identified by Owen et al [6], in which service users rejected the idea of a generic, one-size-fits-all approach for text messages. The consistency of these findings may indicate the importance of these factors in designing SMS text messaging interventions regardless of the cultural contexts of the service users. On the other hand, some factors have not been found in previous studies, such as feeling like a virtual friend. Thus, this finding could be specific to the Chinese cultural background and the population of our service users (adolescents). Adolescence is the period of time when peer relationships and friendships are increasingly established and valued [16]. When facing difficult situations, adolescents are more likely to seek help from their friends rather than from their teachers or parents. This phenomenon might be more common in China, as Chinese parenting has often been described as authoritarian or controlling [17]. In addition, loneliness of adolescents has been associated with a range of mood disorders co-occurring with deliberate self-harm, such as depression [16]. As such, seeking social support from a friend can be especially important and helpful for Chinese adolescents with deliberate self-harm. Having related content in text messaging interventions could also be beneficial for Chinese service users.

Time for Sending Messages

Previous studies of text messaging interventions have not considered the time of the day when the text messages should be delivered [5,7,14]. Our study, therefore, tried to identify particular times of day when the urges of participants to self-harm were the strongest, and situations in which the likelihood of engaging in deliberate self-harm were high. Although some of our participants did not have a fixed time for self-harming, most of them conducted deliberate self-harm during the evening, especially when they were alone. This suggests that more messages could be sent during the evening to reduce the impulses of deliberate self-harm at the right time. In addition, we found that different types of stressful life events such as the divorce of parents and break-up in a romantic relationship could trigger deliberate self-harm among Chinese adolescents. In order to tailor the intervention for individuals, it might be beneficial to send the messages after certain stressful life events have occurred or when they are in crisis.

Limitations

There are a few limitations in the study that warrant discussion. First, the participants in our study were predominantly adolescent girls. As a result, our findings may be more representative of how girls with deliberate self-harm think about text messaging interventions. Although deliberate self-harm shows a higher female-to-male ratio [18], future studies involving a more balanced gender ratio are needed. Second, although 1 participant mentioned that an SMS text messaging intervention does no harm, the interviewer did not specifically ask questions regarding the potential negative effects of such interventions. Nevertheless, this could be an important theme to be mindful of in future research to avoid any adverse outcomes of SMS text messaging interventions. Third, our findings might be generalizable to Chinese adolescents with deliberate self-harm only to a certain extent, as all of our participants had been diagnosed with psychiatric disorders (eg, depression). It would be meaningful to conduct similar research with participants who do not have a clinical diagnosis or who have less severe conditions. Finally, it is worth bearing in mind that China is a big country with 56 ethnic groups, a diverse culture, and a mix of rural and urban areas. This implies that adolescents who were born in different areas in China and belong to different ethnic groups might have different attitudes toward SMS text messaging interventions, which have not been considered in our study.

Conclusions

Our study is the first to investigate the views of adolescents with deliberate self-harm on text messaging interventions. It can be used to inform the development of a culturally tailored SMS text messaging intervention for adolescents with deliberate self-harm. The study has the potential to decrease instances of deliberate self-harm and provide cost-effective, low-intensity, and acceptable support for adolescents with deliberate self-harm who may be reluctant to seek face-to-face psychotherapies.

https://mhealth.jmir.org/2020/6/e16963

JMRI MHEALTH AND UHEALTH  Duan et al
Acknowledgments
This study was supported by the National Natural Science Foundation of China (Numbers: 81974217, 81601197). The funder has no role in the study design, generating or interpreting the results, and publication of the study.

Authors' Contributions
SD and RC designed the study. SD, HW, JQ, and GC conducted all the interviews. SD and JQ conducted the statistical analysis. SD wrote the first draft of the manuscript, and all authors (SD, HW, AW, JQ, GC, YH, YW, JO, and RC) contributed to and have approved the final manuscript.

Conflicts of Interest
None declared.

References

Use of a Machine Learning Program to Correctly Triage Incoming Text Messaging Replies From a Cardiovascular Text–Based Secondary Prevention Program: Feasibility Study

Nicole Lowres¹,², BPhy, PhD; Andrew Duckworth, BSc (Comp Sc Hons 1); Julie Redfern¹,², PhD, BAppSc (Physio Hons 1), BSc; Aravinda Thiagalingam²,³, MB ChB, PhD, FRACP, FCSANZ; Clara K Chow²,³,⁴, MBBS, PhD

¹Heart Research Institute, Sydney, Australia
²Faculty of Medicine and Health, University of Sydney, Sydney, Australia
³Westmead Applied Research Centre, University of Sydney, Sydney, Australia
⁴Department of Cardiology, Westmead Hospital, Sydney, Australia

Corresponding Author:
Nicole Lowres, BPhy, PhD
Heart Research Institute
Building D17, Charles Perkins Centre
Sydney
Australia
Phone: 61 2 8627 5940
Email: nicole.lowres@sydney.edu.au

Abstract

Background: SMS text messaging programs are increasingly being used for secondary prevention, and have been shown to be effective in a number of health conditions including cardiovascular disease. SMS text messaging programs have the potential to increase the reach of an intervention, at a reduced cost, to larger numbers of people who may not access traditional programs. However, patients regularly reply to the SMS text messages, leading to additional staffing requirements to monitor and moderate the patients’ SMS text messaging replies. This additional staff requirement directly impacts the cost-effectiveness and scalability of SMS text messaging interventions.

Objective: This study aimed to test the feasibility and accuracy of developing a machine learning (ML) program to triage SMS text messaging replies (ie, identify which SMS text messaging replies require a health professional review).

Methods: SMS text messaging replies received from 2 clinical trials were manually coded (1) into “Is staff review required?” (binary response of yes/no); and then (2) into 12 general categories. Five ML models (Naïve Bayes, OneVsRest, Random Forest Decision Trees, Gradient Boosted Trees, and Multilayer Perceptron) and an ensemble model were tested. For each model run, data were randomly allocated into training set (2183/3118, 70.01%) and test set (935/3118, 29.98%). Accuracy for the yes/no classification was calculated using area under the receiver operating characteristics curve (AUC), false positives, and false negatives. Accuracy for classification into 12 categories was compared using multiclass classification evaluators.

Results: A manual review of 3118 SMS text messaging replies showed that 22.00% (686/3118) required staff review. For determining need for staff review, the Multilayer Perceptron model had highest accuracy (AUC 0.86; 4.85% false negatives; and 4.63% false positives); with addition of heuristics (specified keywords) fewer false negatives were identified (3.19%), with small increase in false positives (7.66%) and AUC 0.79. Application of this model would result in 26.7% of SMS text messaging replies requiring review (true + false positives). The ensemble model produced the lowest false negatives (1.43%) at the expense of higher false positives (16.19%). OneVsRest was the most accurate (72.3%) for the 12-category classification.

Conclusions: The ML program has high sensitivity for identifying the SMS text messaging replies requiring staff input; however, future research is required to validate the models against larger data sets. Incorporation of an ML program to review SMS text messaging replies could significantly reduce staff workload, as staff would not have to review all incoming SMS text messages. This could lead to substantial improvements in cost-effectiveness, scalability, and capacity of SMS text messaging–based interventions.

(JMIR Mhealth Uhealth 2020;8(6):e19200) doi:10.2196/19200
**KEYWORDS**

eHealth; machine learning, secondary prevention, SMS text messaging, cardiovascular, mHealth, digital health, mobile phone

**Introduction**

Cardiovascular disease (CVD) is the leading cause of death globally [1]. Secondary prevention interventions, such as cardiac rehabilitation, are very effective at reducing CVD and preventing cardiovascular events [2,3]. However, access to traditional cardiac rehabilitation and secondary prevention is limited due to many well-described barriers [4,5]. As a result, alternative methods of secondary prevention, including electronic health (eHealth) and mobile health (mHealth), are emerging with demonstrated effectiveness [6]. Because 88% of Australians (aged 18-75 years) now own a smartphone [7], the potential reach of mHealth programs to deliver CVD secondary prevention is significant. Smartphone mHealth interventions are shown to be an acceptable and feasible method of health care delivery [8,9]. Specifically, SMS text messaging programs have been shown to be effective for cardiac risk factor reduction [10-12], medication adherence [13,14], weight loss [15], and physical activity [16,17]. However, one study showed limited maintenance of behavior changes once the SMS text messaging intervention ceased [17]. SMS text messaging programs have also been utilized effectively in many other medical settings including diabetes [18], maternal health [19], hepatitis B [20], and smoking cessation [21].

SMS text messaging programs can be set up as either a one- or a two-way messaging communication system. The TEXTME program was an automated semipersonalized SMS text messaging program to support lifestyle change for people with coronary heart disease, which was originally delivered as a one-way communication system [12]. TEXTME sent participants 4 SMS text messages per week for 6 months, and demonstrated significant improvements in cholesterol, blood pressure, BMI, physical activity, and smoking reduction, compared with usual care [12]. In addition, participants were highly satisfied and engaged with the program, reporting numerous additional psychosocial benefits [12,22]. However, even in this one-way SMS text messaging program, where participants were specifically instructed not to respond, 32.9% (116/352) replied to the SMS text messages, with many replying multiple times [22]. The cost of delivering automated SMS text messages is relatively inexpensive (approximately AUD 13/person [US $8.60/person] for the entire program) [12]. However, when participants respond to SMS text messages there are various staffing and logistical requirements and a trained health professional is required to moderate replies. In TEXTME many replies were statements of thanks, but some SMS text messages require a professional response, particularly in the case of medical distress [22], or if participants wish to withdraw from the program.

Machine learning (ML) programs have commonly been used in health care to predict outcomes [23,24], but in other applications they have been used to classify text [25,26]. It is therefore possible that a computerized ML program could be developed and trained to accurately identify which incoming SMS text messages require review by a health professional. If successful, an ML program could significantly reduce the health professional’s workload. Thus, the cost of running an SMS text messaging–based intervention would be significantly reduced, and the scalability and capacity of the program would improve substantially. Therefore, this study aimed to test the feasibility of developing an ML program to triage and identify SMS text messages requiring a health professional’s review. Specifically, the primary aim was to determine the accuracy of the ML program to sort which SMS text messaging replies require a staff member review (ie, binary classification of yes/no); and the secondary aim was to assess the accuracy of the ML program to classify each SMS text messaging reply into overall categories or themes (ie, multiclass classification).

**Methods**

**Study Population**

The SMS text messaging replies used in this analysis originated from the TEXTME and TEXTMEDS programs [12,27]. TEXTME was a 6-month program of SMS text messages for patients with coronary heart disease, where patients were explicitly instructed not to respond to SMS text messages (ACTRN12611000161921) [12]. TEXTMEDS was a 12-month program for patients with acute coronary syndrome, with a more interactive two-way SMS text messaging protocol (ACTRN12613000793718) [27]. Both studies recruited patients during hospitalization. The demographics of the TEXTME patients have previously been published [12], which included the following: mean age 57.9 (SD 9.1) years; 81.5% (287/352) men; and education level of 11 years (interquartile range 9-13).

**Manual Coding of SMS Text Messaging Replies**

In this study, de-identified SMS text messaging replies from the TEXTME and TEXTMEDS studies were extracted into a database. SMS text messages were entered into the database exactly as they were received, with no correction to spelling, grammar, or punctuation. The TEXTME and TEXTMEDS programs had created 22 categories of SMS text messaging replies for administrative and management purposes (eg, according to content, administrative request, or general commentary; Multimedia Appendix 1).

Two experienced cardiac health professionals (NL and Anu Indrawansa, Westmead Hospital, Sydney, Australia) independently reviewed and coded each SMS text message into the 22 categories according to the theme of the message. For the SMS text messages that expressed more than 1 theme, categorization was done according to the main theme or the theme that required a staff member review/action. For SMS text messages where a consensus on coding could not be reached, a third person (AD) was consulted. Using an iterative process, these 22 categories were condensed by NL and AD into 12 categories (Table 1), with consideration given to the type of response required for the SMS text message. Finally, each SMS text message was designated into one of two groups, depending...
on whether it required a staff member review or not (ie, yes or no).

**Machine Learning Model Development**

Five different ML models and an ensemble model were tested using Apache Spark (version 2.3.4) and the models, and their associated parameters or weights or both, were saved for further analysis. First, we assessed 4 models which are considered good models for classification purposes: Naïve Bayes, OneVsRest, Random Forest decision trees, and Gradient Boosted Trees. These models use a combination of different statistical data analytics and regression tools to assess the data and determine the likelihood of the correct response. Second, we assessed a convolutional neural network classification approach: Multilayer Perceptron. Neural networks use layers of iterative feedback cycles to calculate the outcome, and are commonly used for natural language translation in systems such as Google Translate, and for speech and image recognition. Lastly, we implemented a technique that combined the results from different models together as an *Ensemble model*. We evaluated the effect of using heuristics (ie, selected keywords) in each model, by programming and testing each model with and without the inclusion of heuristics. We also evaluated the addition of *natural language programming* which interprets and determines the intended context of written statements.

Each time a model was run, the data set was randomly divided into a training set (2183/3118, 70.01%) and a test set (935/3118, 29.98%), thus creating a unique allocation of SMS text messages to the training and test data each time. Each model was assessed and compared for determining if health professional review was required (ie, binary variable of yes/no: *Figure 1*). In addition, we assessed and compared each model that supported multiclass classification (Gradient Boosted Trees only supports binary classification) for its ability to correctly classify the SMS text messages into the 12 categorical variables as listed in *Table 1*. Each model was run 5 times on different random splits of the training and test data sets to validate the accuracy of the results.

### Table 1. Manual classification of SMS text messaging replies (N=3118)

<table>
<thead>
<tr>
<th>Classification category</th>
<th>Health professional review required</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes\textsuperscript{a}, n (%)</td>
<td>No\textsuperscript{b}, n (%)</td>
</tr>
<tr>
<td>Health question/concern</td>
<td>383 (12.28)</td>
<td>–</td>
</tr>
<tr>
<td>Administrative request</td>
<td>155 (4.97)</td>
<td>–</td>
</tr>
<tr>
<td>Request to STOP/pause</td>
<td>107 (3.43)</td>
<td>–</td>
</tr>
<tr>
<td>Ceased smoking</td>
<td>24 (0.76)</td>
<td>–</td>
</tr>
<tr>
<td>SMS text message not delivered</td>
<td>14 (0.44)</td>
<td>–</td>
</tr>
<tr>
<td>Timely response required</td>
<td>4 (0.12)</td>
<td>–</td>
</tr>
<tr>
<td>General statement</td>
<td>–</td>
<td>1132 (36.30)</td>
</tr>
<tr>
<td>Statement of thanks</td>
<td>–</td>
<td>789 (25.30)</td>
</tr>
<tr>
<td>Reporting good health</td>
<td>–</td>
<td>216 (6.92)</td>
</tr>
<tr>
<td>Blank SMS text message</td>
<td>–</td>
<td>191 (6.12)</td>
</tr>
<tr>
<td>Positive emoticon</td>
<td>–</td>
<td>74 (2.37)</td>
</tr>
<tr>
<td>Unrelated/accidental</td>
<td>–</td>
<td>29 (0.93)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Total: 687/3118 (22.03%).

\textsuperscript{b}Total: 2431/3118 (77.96%).
Figure 1. Machine learning model development and testing.

Statistical Analysis and Comparison of Models

For the primary outcome of determining which SMS text messaging replies require a staff member review (binary outcome of yes/no), a binary classification evaluator was used to calculate accuracy using the area under the receiver operating characteristics curve (AUC), the rate of true and false positives, and the rate of true and false negatives. For the secondary outcome of the accuracy of the ML program to classify the SMS text messaging replies into the 12 categories, a multiclass classification evaluator was used to compare each model for accuracy, precision, recall, and F1 score.

Results

Manual Classification of SMS Text Messaging Replies

A manual review of 3118 SMS text messaging replies was performed. The manual classification identified 12 broad categories of replies (Table 1). Only 22.03% (687/3118) of all SMS text messages required someone to review and reply to the SMS text messages, and only 0.12% (4/3118) SMS text messages were deemed to warrant a timely response within 24-48 h; however no SMS text messages were considered urgent (Table 1).

Machine Learning Model Development

Natural language programming was evaluated as an option to improve the ML model’s ability to understand the sentiment behind each SMS text message. Ultimately, natural language programming was not included in any of the models as challenges arose in relation to the SMS text messaging content, due to the frequent use of nonstandard English grammar, nonstandard abbreviations, and spelling errors within the SMS text messages.

Accuracy for Determining If Staff Review Is Required

The results for each model, with and without the inclusion of heuristics, are outlined in Figure 2 and Multimedia Appendix 2. The inclusion of heuristics reduced the number of false negatives in each of the models, at the expense of an increase in the number of false positives. From the individual models, Naïve Bayes with heuristics produced the lowest false negatives (1.98%) and the highest false positives (14.54%). The Random Forest decision trees model resulted in the lowest number of false positives (1.87%), but had the highest rate of false negatives (13.44%).

The Multilayer Perceptron model correctly identified if staff review was required (yes/no) in 90.52% of cases (4.85% false negatives and 4.63% false positives from the test data sample). With addition of heuristics to the Multilayer Perceptron model, fewer false negatives were identified (3.19%) with a small increase in false positives (7.71%). Utilizing the Multilayer Perceptron model with heuristics, staff would have to review only 26.65% of all SMS text messaging replies (ie, 18.94% true positives + 7.71% false positives; Figure 3).
However, with the aim of reducing the false negatives to the lowest possible number, the ensemble model was found to be superior (1.43% false negatives and 16.2% false positives), giving a sensitivity of 93.5% for identifying the SMS text messages requiring review, and a specificity of 81.3%. Using the ensemble model with heuristics, the results indicated that health professionals would have to review 36.9% (ie, 20.7%
true positives + 16.2% false positives) of all incoming SMS text messaging replies (Figure 3).

Examples of SMS text messages that were incorrectly coded by the Multilayer Perceptron model as “Yes – review required”:

I am now careful with my choices when i purchase meats etc bacon is not eaten very often like it used to be

Pill are not a problem for me. To this point limited to no side effects. Just got to keep up with exercise regime.

Examples of SMS text messages that were incorrectly coded by the Multilayer Perceptron model as “No review required”:

Hi everything is excellent. Eating well exercising everyday taking medicine. You could probably take me of list as everything you suggest has already been implemented from day one of recovery. Thanks for your input. Regards

Just a query, I walk for 1/2 each day, is this sufficient?

Accuracy to Identify the Priority SMS Text Messages

Importantly, the model correctly identified the 4 priority SMS text messages that required a timely review. The model missed only 1 of 107 SMS text messages where the participant was requesting to opt out of the program or pause the SMS text messages. The only SMS text message that was missed read “no more texts now thanks.”

Accuracy for Classification Into 12 Categories

For classification into the 12 categories, OneVsRest and the Multilayer Perceptron models produced the best results. OneVsRest had slightly higher accuracy of 0.723, with precision 0.723, recall 0.723, and F1 score 0.719. For the Multilayer Perceptron model, accuracy was 0.717, precision 0.735, recall 0.717, and F1 score 0.723.

Discussion

Principal Findings

Our study demonstrates the feasibility and accuracy of using an ML program as a triage system to sort which incoming SMS text messaging replies require health professional review, in a cardiovascular secondary prevention SMS text messaging intervention setting. Each of the ML models tested varied in their sensitivity and specificity for classifying the SMS text messages, and there appeared to be a trade-off between false negatives and false positives. From the individual models, Naïve Bayes produced the lowest false-negative rate, and Random Forest decision trees produced the lowest false-positive rate. In our study, the ensemble model (with all models combined) was the most accurate at identifying the SMS text messages needing review, with only 1.43% of the SMS text messages being false negatives. If this ensemble model were utilized as a triage for incoming SMS text messages, health professionals would have to review about 37% of all the incoming SMS text messaging replies. For sorting the SMS text messages into 12 categories (according to the theme of the SMS text message), the accuracy was only moderate (0.723), indicating further work is needed to create suitable categories and to train the model with larger data sets and examples.

To our knowledge this is the first study in a health setting to use an ML program to classify (ie, triage) incoming SMS text messages. ML has been widely validated and used for the purpose of filtering SPAM SMS text messages, classifying SMS text messages into a binary outcome (SPAM or not SPAM) which is similar to our study [28,29]. These studies on SPAM SMS text messages have employed different approaches for reviewing content, filtering, and feature extraction; however, they have produced similar accuracy results to our study [28,29]. The use of ML SPAM filters within the operating systems of smartphones is now commonplace, with early adoption dating back to Apple’s iOS 11 in 2017.

Within the health industry, ML algorithms are common in technology such as electrocardiogram analysis; however, the use of ML to interpret medical text or within patient–provider interactions has not been adopted in everyday practice. There may be larger ethical barriers to overcome in the adoption of ML programs if they are used to assist health care decisions and advice, due to the duty of care to our patients and concern over the consequences if the ML program misses a critical medical alert. Research using ML in the health setting has predominantly focused on extracting information from fields in health and medical records to predict the risk of an outcome such as mortality [30,31], emergency re-admissions [32], and myopia development [33]. It is difficult to compare accuracy across these studies as they measured different outcomes and reported accuracy using different metrics, however they generally performed well, with accuracy examples of AUC > 0.82 for predicting mortality after echocardiography [30]; and AUC range of 0.87-0.98 for predicting myopia development [33]. There is one study which has used ML as a triage for patients with congenital heart disease to estimate the need for a multidisciplinary review [34]. With all these studies, the ML models were trained to review fields in the medical records to determine the likelihood of the outcome, and therefore did not have to interpret the meaning or sentiment of the text in the medical notes [34].

More recently, ML has successfully been applied to more unstructured written health notes and dialogues [35,36]. One study was able to phenotype depression using unstructured notes in the medical records with a sensitivity of 93.5% and specificity of 68% [35]. Natural language processing has also been used to review secure online discussions between patients and health professionals, in which linguistic features were assessed to predict health literacy levels with an accuracy of 60.55% (C-statistic 0.63) [36]. Common to these studies, there was one ML model which performed better than the other models, and the model varied across the studies dependent on the primary data and outcome they were measuring, and this is also reflected in our results.

Clinical Implications

Although ML research in health care is still evolving, the possible breadth of using ML is promising. The most valuable feature of the ML program in our study was the ability to...
identify the SMS text messages that were needed to be reviewed by a staff member. Therefore, we were interested in identifying the model with the lowest false-negative rate. The AUC was not the best statistic for assessing the model performance, as for each small decrease in false negatives, there was a much larger increase in false positives, and thus a reduction in AUC.

If the ensemble model, with the lowest false negatives, was implemented in a secondary prevention SMS text messaging program, the minimal number of SMS text messages will be missed, but this would be at the expense of staff needing to review more overall SMS text messages (due to the higher false positives). However, even with reviewing additional false positives, staff would only need to review approximately 37% of all incoming SMS text messages, which will significantly reduce the workload of the health professionals running the program, permitting more efficient use of their time. Although it is thought that SMS text messaging interventions may be cost-effective [37], the cost-effectiveness of this solution would need to be evaluated prior to implementation.

Furthermore, prior to implementation it is also important to ensure that the model is capturing all of the most important SMS text messages requiring action from the health professional, and that there is no systemic misclassification. This would need to be tested with larger databases, with careful monitoring of the high-risk categories, to ensure the model is doing what we want it to do at scale. The use of heuristics in the model provides an additional safeguard to ensure that SMS text messages with certain keywords are reviewed. If participants are provided with specific words to opt out of SMS text messaging programs, such as the keyword STOP, then heuristics are even more effective at identifying all participants trying to opt out of the program.

Once implemented, the ML program would run alongside the SMS text messaging program, and would continue to be trained as the database grows, thereby improving the accuracy of the ML algorithms. To further improve the certainty of SMS text message classification, it would be possible to program BOT-like features into the model to automatically reply to the sender if clarification of the need to respond is required (eg, “Hi...Would you like me to contact you to discuss this?”). Furthermore, if the accuracy of the ML model could be improved for the classification of SMS text message into more specific categories, it may be possible to consider developing an automated system that sends an appropriate motivational response for the SMS text messages that do not require a staff member response.

**Strengths and Limitations**

This is the first study in a health setting to explore the use of ML for classification of health-related SMS text messages. It assesses and compares the accuracy of 7 different ML models, and employs a robust method for verification of results using 5 different splits of the training and test data sets. In all triage systems that require a human to decide on the priority and need for medical care, there is a certain element of subjective bias in the decision-making process. In our study where health professionals classified the incoming SMS text messages into whether or not a review was required, an element of subjectivity in that decision remained. This decision making is also complicated by many of the SMS text messages expressing one or more sentiments, and a subjective decision being made as to the most important sentiment to classify the SMS text messages. As there is no gold standard to define whether an SMS text message needs to be reviewed or not, it is possible that health professionals with different areas or levels of experience may manually classify the incoming SMS text messages differently. Our study demonstrates a proof-of-concept, as we tested a limited number of SMS text messaging replies (n=3118); however, to generalize the results beyond our study population, future work is required to validate the model against larger data sets, and on data sets that are coded by different health professionals.

**Conclusions**

In our feasibility study, the ML program displayed high sensitivity in identifying the SMS text messaging replies that require health professional input. There is a low false-negative rate, indicating few messages which need a response would be missed. Introduction of the ML program to SMS text messaging--based programs could therefore significantly reduce the need for health professionals to review every incoming SMS text message. This could lead to substantial improvements in scalability and capacity of SMS text messaging--based programs. The future implications for this technology are vast, including potential utilization in other interactive mHealth interfaces and cardiovascular health apps.

**Acknowledgments**

We gratefully acknowledge Ms Anu Indrawansa for her invaluable assistance in the running of the trials, database management, and coding of the SMS text messaging databases. NL is funded by an NSW Health Early Career Fellowship (H16/52168). JR is funded by an NHMRC Career Development Fellowship (APP1143538). CKC is funded by an NSW Health Clinician scientist grant. The TEXTME study was supported by peer-reviewed grants from the National Heart Foundation of Australia Grant-in-Aid (G1055110) and a BUPA Foundation Grant. The TEXTMEDS study was funded by a National Health and Medical Research Council (NHMRC) Project grant (ID APP1042290). This work was presented in part as a poster presentation at the European Society of Cardiology Conference, Paris, in August 2019.

**Authors’ Contributions**

All authors contributed to the conception and study design of the work. CKC, JR, and AT contributed to acquisition of the raw data. NL and AD contributed to the analysis and interpretation of data for the work. NL drafted the manuscript. All authors...
critically revised the manuscript and gave final approval and agree to be accountable for all aspects of work ensuring integrity and accuracy.

**Conflicts of Interest**

AD is an employee of Sandstone Technology: work on this study is not related to Sandstone Technology and was performed external to his employment. All other authors declare no conflicts of interest.

**Multimedia Appendix 1**

Original Categorisation used in TEXTME and TEXTMEDS study.

**Multimedia Appendix 2**

Results for identification of texts requiring review.

**References**


21. Gram IT, Larbi D, Wangberg SC. Comparing the efficacy of an identical, tailored smoking cessation intervention delivered by mobile text messaging versus email: randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 27;7(9):e12137 [FREE Full text] [doi: 10.2196/12137]


Abbreviations

AUC: area under the receiver operating characteristics curve
CVD: cardiovascular disease
ML: machine learning

©Nicole Lowres, Andrew Duckworth, Julie Redfern, Aravinda Thiagalingam, Clara K Chow. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 16.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Supplemental Text Message Support With the National Diabetes Prevention Program: Pragmatic Comparative Effectiveness Trial

Natalie D Ritchie¹, PhD; Silvia Gutiérrez-Raghunath¹, MBA; Michael Josh Durfee¹, MSPH; Henry Fischer¹, MD Denver Health, Denver, CO, United States

Corresponding Author: Natalie D Ritchie, PhD Denver Health 601 Broadway Denver, CO, 80204 United States Phone: 1 3036022188 Email: natalie.ritchie@dhha.org

Abstract

Background: The evidence-based National Diabetes Prevention Program (NDPP) is now widely disseminated, yet strategies to increase its effectiveness are needed, especially for underserved populations. The yearlong program promotes lifestyle changes for weight loss and can be offered in-person, online, via distance learning, or a combination of modalities. Less is known about which delivery features are optimal and may help address disparities in outcomes for subgroups. We previously demonstrated the efficacy of a stand-alone text messaging intervention based on the NDPP (SMS4PreDM) in a randomized controlled trial in a safety net health care system. Upon broader dissemination, we then showed that SMS4PreDM demonstrated high retention and modest weight loss at a relatively low cost, suggesting the potential to improve in-person NDPP delivery.

Objective: In this study, we aim to compare the effectiveness of in-person NDPP classes with and without supplementary SMS4PreDM on attendance and weight loss outcomes to determine whether text messaging can enhance in-person NDPP delivery for a safety net patient population.

Methods: From 2015 to 2017, patients with diabetes risks were identified primarily from provider referrals and enrolled in NDPP classes, SMS4PreDM, or both per their preference and availability. Participants naturally formed three groups: in-person NDPP with SMS4PreDM (n=236), in-person NDPP alone (n=252), and SMS4PreDM alone (n=285). This analysis compares the first two groups to evaluate whether supplemental text messaging may improve in-person NDPP outcomes. Outcomes for SMS4PreDM-only participants were previously reported. NDPP classes followed standard delivery guidelines, including weekly-to-monthly classes over a year. SMS4PreDM delivery included messages promoting lifestyle change and modest weight loss, sent 6 days per week for 12 months. Differences in characteristics between intervention groups were assessed using chi-square and t tests. Differences in NDPP attendance and weight loss outcomes were analyzed with multivariable linear and logistic regressions.

Results: The mean age was 50.4 years (SD 13.9). Out of a total of 488 participants, 76.2% (n=372) were female and 59.0% (n=288) were Hispanic. An additional 17.2% (n=84) were non-Hispanic white and 12.9% (n=63) were non-Hispanic black. A total of 48.4% (n=236) of participants elected to receive supplemental text message support in addition to NDPP classes. Participants who chose supplemental text message support were on average 5.7 (SD 1.2) years younger ($P < .001$) than the 252 participants who preferred in-person classes alone. Relatively more women and Hispanic individuals enrolled in the NDPP with supplemental text messages than in NDPP classes alone, 83.9% (n=198) vs 69.0% (n=174, $P < .001$) and 68.6% (n=162) vs 50.0% (n=126, $P = .001$), respectively. Attendance and weight loss outcomes were comparable between groups.

Conclusions: Despite its appeal among priority populations, supplemental text messaging did not significantly increase attendance and weight loss for the in-person NDPP. Further research is needed to identify optimal strategies to improve the effectiveness of the NDPP.

(JMIR Mhealth Uhealth 2020;8(6):e15478) doi:10.2196/15478

KEYWORDS
eHealth; prediabetes; texting; weight loss
Introduction

The Diabetes Prevention Program was a successful clinical trial to prevent type 2 diabetes, showing intensive lifestyle support reduced incidence by 58% [1], with positive effects lasting long-term [2,3]. The lifestyle intervention was translated into the National Diabetes Prevention Program (NDPP) and has been widely disseminated since 2012 [4]. Per delivery standards established by the Centers for Disease Control and Prevention (CDC), the yearlong program can be offered in-person, online, via distance learning, or a combination of modalities [5]. Based on the original trial, the NDPP promotes ≥5% weight loss through diet and physical activity to reduce diabetes risk [5]. Strategies to optimize the NDPP dissemination are needed to achieve this weight loss goal in real-world practice.

For the in-person NDPP, a national study found a promising 4.2% mean weight loss, but early dropout is problematic and limits weight loss for many participants [4]. Further, there are concerning disparities in outcomes for subgroups including racial or ethnic minorities, low-income non-Hispanic whites, and younger adults, who achieve about half of the weight loss of their respective counterparts [4,6,7]. Disparities in weight loss have largely been attributed to limited attendance [4,6].

Virtually delivered programming may be more convenient than in-person classes by removing barriers like lack of transportation and being more attractive to younger individuals, yet shows comparable overall weight loss (mean 4.3%) and is understudied in diverse groups [8]. Less is known about which virtual delivery methods may enhance the NDPP outcomes [9], which is important given the performance-based payment models for the program [10,11]. Supplemeting virtually delivered interventions with in-person coaching has been shown to support modest improvements in weight loss (mean 4.6%) [8]. Conversely, whether concurrent virtually delivered education and support could improve the effectiveness of the in-person NDPP appears unknown.

We previously demonstrated the efficacy of a stand-alone text messaging intervention based on the NDPP curriculum (SMS4PreDM) in a randomized controlled trial of 163 patients in an urban safety net health care system (Denver Health) [12]. In a pragmatic effectiveness study of wider SMS4PreDM dissemination at Denver Health, we then showed that stand-alone SMS4PreDM was relatively low in cost to deliver (US $100.92 per each of the 285 participants) and demonstrated high retention, albeit with modest weight loss compared to 1233 observation-only controls [13]. Now, to determine whether text messaging can enhance in-person NDPP delivery, we compare effectiveness of the in-person NDPP with and without supplementary SMS4PreDM on attendance and weight loss outcomes.

Methods

Denver Health serves one-third of Denver, Colorado residents across its network of community- and school-based clinics, specialty care centers, and a Level 1 hospital. The majority of patients are low-income and of minority racial or ethnic backgrounds [14]. From October 2015 to October 2017, Denver Health offered the in-person NDPP, SMS4PreDM, or both to patients at risk for diabetes. New NDPP classes and concurrent SMS4PreDM programming began approximately each quarter, with advance announcements to Denver Health providers about opportunities to refer eligible patients. Per CDC criteria, eligible participants included adults with a BMI ≥24 (≥22 if Asian) and with prediabetes (ie, hemoglobin A1c 5.7-6.4), a former diagnosis of gestational diabetes, or a positive score on a diabetes risk questionnaire [15]. Lifestyle coaches then conducted outreach calls to verify eligibility, preference, and availability to enroll in upcoming NDPP classes, SMS4PreDM, or both. As of September 2016, patients who initially selected in-person NDPP classes (with or without SMS4PreDM) were encouraged to first attend an in-person orientation session designed to increase engagement in the NDPP [16]. At these orientation sessions, patients who had not initially enrolled in SMS4PreDM could also opt in at this time. There were no fees or financial incentives to participate in these risk reduction programs.

Participants naturally formed three groups: those receiving (1) in-person NDPP classes with SMS4PreDM (n=236), (2) NDPP classes alone (n=252), and (3) SMS4PreDM alone (n=285). As already noted, outcomes for the SMS4PreDM-only group were previously reported [13]. This study focuses on determining whether supplemental text message support may improve delivery of in-person NDPP classes, thus comparing the first two groups of participants receiving NDPP classes plus SMS4PreDM vs NDPP classes alone (N=488).

The in-person NDPP was conducted following guidelines established by the CDC [15]. Participants were offered a total of 22-25 sessions over 1 year (depending on scheduling). The classes complied with standards to provide at least 16 hour-long group sessions in months 1-6 and a minimum of 6 sessions in months 7-12 [15]. Trained, bilingual lay health educators served as lifestyle coaches and led sessions in English or Spanish. Classes were held at Denver Health’s community clinics and main campus locations. SMS4PreDM participants received messages promoting lifestyle change and modest weight loss, delivered 6 days per week for 12 months, following our published SMS4PreDM methodology [12]. Content followed the NDPP session schedule for concordance with the topics of in-person classes. SMS4PreDM was also available in English and Spanish.

Demographic characteristics and data on diabetes risks were collected from electronic health records and self-report as needed. Attendance outcomes were the percentage of NDPP sessions attended and days of NDPP attendance (1-365), per previous findings that intensity and duration of participation are key drivers of weight loss in the NDPP [4]. Weight loss outcomes were total percent of body weight lost and achieving ≥5% weight loss in the NDPP, based on weights recorded at the first and last sessions attended. Body weight was measured at each NDPP session on a high-capacity, medical-quality scale. Participants were encouraged to wear consistent attire for weight measurements (eg, removing outerwear). Data collection was completed in September 2018.

http://mhealth.jmir.org/2020/6/e15478/
Differences in characteristics between intervention groups were assessed using chi-square and t tests. Differences in NDPP attendance and weight loss outcomes were analyzed with multivariable linear and logistic regression. Covariates included age, gender, and race or ethnicity. We further controlled for effects of attending orientation sessions, as they were previously found to improve the NDPP outcomes [16]. Post hoc analyses explored language as a moderating variable among Hispanic participants, given that Spanish-speaking (vs English-speaking) participants previously had more weight loss in SMS4PreDM alone [12,13] but not in NDPP classes alone [6]. Analyses were completed using SPSS version 22 (IBM Corp). The Colorado Multiple Institutional Review Board approved this program evaluation project, which was not registered as a clinical trial.

**Results**

The majority of the 488 participants were female (n=372, 76.2%) and Hispanic (n=288, 59.0%), among whom 67.4% (n=194) preferred Spanish. An additional 17.2% (n=84) were non-Hispanic white and 12.9% (n=63) were non-Hispanic black. Mean age was 50.4 years (SD 13.9). Average BMI at enrollment was 34.8 (SD 7.7). Almost half (n=236, 48.4%) of participants elected to receive supplemental text messaging support while attending in-person NDPP classes, and 252 received NDPP class only.

Relatively more women and Hispanic individuals enrolled in the NDPP with supplemental text messages than in NDPP classes alone, 83.9% (n=198) vs 69.0% (n=174; P<.001) and 68.6% (n=162) vs 50.0% (n=126; P=.001), respectively. Participants who chose supplemental text message support were on average 5.7 (SD 1.2) years younger (P<.001) than those preferring in-person classes alone. There were no other significant differences in demographic characteristics or starting BMI between groups. Over one-third (n=193, 39.5%) of participants in this analysis first attended an orientation session. Those who joined NDPP classes after recruitment from an orientation session were more likely to choose supplemental text messages than patients enrolled via outreach calls alone (n=150, 77.7% vs n=86, 29.2%, P<.001).

There were no significant differences in attendance and weight loss outcomes between NDPP participants who received supplemental text messages and those who did not (Table 1). Language did not moderate intervention effectiveness among Hispanic participants, whether for attendance or weight loss outcomes. For example, achieving ≥5% weight loss was not significantly modified by language (P=.60).

There were no significant differences in attendance and weight loss outcomes between NDPP participants who received supplemental text messages and those who did not (Table 1). Language did not moderate intervention effectiveness among Hispanic participants, whether for attendance or weight loss outcomes. For example, achieving ≥5% weight loss was not significantly modified by language (P=.60).

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>NDPP with supplemental text message support (n=236), mean (SE)</th>
<th>NDPP without supplemental text message support (n=252), mean (SE)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDPP sessions attended (%)</td>
<td>38.1 (0.0)</td>
<td>36.8 (0.0)</td>
<td>.68</td>
</tr>
<tr>
<td>Number of days in NDPP (1-365)</td>
<td>146.0 (9.8)</td>
<td>134.8 (9.9)</td>
<td>.46</td>
</tr>
<tr>
<td>Weight loss in NDPP (%)</td>
<td>2.1 (0.3)</td>
<td>1.7 (0.3)</td>
<td>.37</td>
</tr>
<tr>
<td>Achieved ≥5% weight loss in NDPP (%)</td>
<td>44 (18.6)</td>
<td>37 (14.2)</td>
<td>.84</td>
</tr>
</tbody>
</table>

aData are presented as adjusted mean and SE of the mean with multivariable linear regression P values for continuous variables and unadjusted frequency (%) with multivariable logistic regression P values for categorical variables. Adjusted odds ratio for achieving ≥5% weight loss in NDPP with supplemental text message support is 1.065 (95% CI 0.584-1.944, P=.84). Covariates include age, gender, race or ethnicity, and completion of orientation session.

| NDPP: National Diabetes Prevention Program. |

**Discussion**

In a safety net health care system, nearly half of patients with diabetes risks who joined in-person NDPP classes elected to also receive supplemental text message support. Supplemental text message support seemed to appeal relatively more to women, Hispanic individuals, and younger participants than other groups. Despite its appeal among priority populations, supplemental text messaging did not significantly increase attendance and weight loss in the NDPP. The extra costs of supplemental text message support may be unwarranted without sufficiently improved risk reduction. Nonetheless, supplemental text message support did not appear detrimental; participants receiving both in-person and virtual content attended classes equally, despite remote access to content.

Overall, weight loss was limited for in-person NDPP participants in this study, both with and without supplemental text message support (2.1% and 1.7% averages, respectively), when contrasted with the national average of 4.2% weight loss [4]. This may reflect overall challenges of serving a safety net patient population and highlights that additional improvements are needed. Further study to increase effectiveness of text messaging may be indicated given that retention is problematic in yearlong in-person classes, contrasted with high retention previously shown for SMS4PreDM participants. Specifically, a national evaluation showed that the majority of participants complete less than half of the in-person NDPP [4], whereas 91% of participants completed the standalone SMS4PreDM intervention [13]. Standalone SMS4PreDM was associated with only minimal weight loss (≤1%) in real-world practice [13], yet even small amounts of weight loss can be clinically meaningful—each kilogram lost is associated with a 16% reduction in diabetes incidence [17]. Future research may be merited to determine whether participants who discontinue in-person NDPP classes may then benefit from acceptable alternatives such as virtually delivered content. At the same time, developing other strategies to improve the NDPP outcomes appears needed, such as motivational enhancements, identifying and removing...
participation barriers, and addressing social determinants of health. Limitations include the nonrandomization in this pragmatic study, which likely contributed to demographic differences between comparator groups; although, analyses controlled for these factors. Additional orientation sessions may have better conveyed the opportunity to receive supplemental text message support than outreach calls alone did; although, we also controlled for attendance to these introductory sessions. Participants who discontinued in-person NDPP classes had continued access to SMS4PreDM content, and the extent to which the continued supplemental text message support may have benefited participants is unknown. Further, we were unable to distinguish participants who meaningfully engaged in supplemental text messaging.

In conclusion, findings suggest that, although many patients at high risk for type 2 diabetes select supplemental text message support, this resource does not increase effectiveness of the in-person NDPP. As such, this study may contribute to knowledge about which dissemination features are important or not for delivery of the NDPP.

Acknowledgments

The National Diabetes Prevention Program at Denver Health was funded largely by the Amendment 35 Cancer, Cardiovascular Disease and Pulmonary Disease Grant Program administered by the Colorado Department of Public Health and Environment, as well as an award from America’s Health Insurance Plans in partnership with the Centers for Disease Control and Prevention. Additional support was provided by Denver Health. Authors report no other conflicts of interest. The contents of this publication are solely the responsibility of the authors and do not represent official views of these organizations.

Authors’ Contributions

Author NDR is principally responsible for the presented study, including study design, data access, and the decision to submit and publish the manuscript. Authors NDR, SGR, MJD, and HF conceived the research. NDR conducted the data analysis and wrote the manuscript. All authors critically reviewed the manuscript, and all authors read and approved the final submitted version.

Conflicts of Interest

None declared.

References


Abbreviations

CDC: Centers for Disease Control and Prevention
NDPP: National Diabetes Prevention Program

©Natalie D Ritchie, Silvia Gutiérrez-Raghunath, Michael Josh Durfee, Henry Fischer. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 18.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Perceptions About Mindfulness and Text Messaging for Smoking Cessation in Vietnam: Results From a Qualitative Study

Vuong Van Do, MPH; Claire Adams Spears, PhD; Hoang Van Minh, PhD; Jidong Huang, PhD; Pamela Buffington Redmon, MPH; Nguyen Xuan Long, PhD; Michael Paul Eriksen, ScD

1Center for Population Health Sciences, Hanoi University of Public Health, Hanoi, Vietnam
2School of Public Health, Georgia State University, Atlanta, GA, United States
3Emory Global Health Institute, Emory University, Atlanta, GA, United States
4Vietnam Association of Psychology, Hanoi, Vietnam

Corresponding Author:
Vuong Van Do, MPH
Center for Population Health Sciences
Hanoi University of Public Health
1A Duc Thang, Bac Tu Liem
Hanoi, 100000
Vietnam
Phone: 1 7702352346
Email: dvv@huph.edu.vn

Abstract

Background: With 15.6 million smokers, Vietnam is one of the top 10 largest cigarette-consuming countries in the world. Unfortunately, smoking cessation programs are still scarce in Vietnam. Mindfulness-based and text messaging–based interventions have been increasingly used in smoking cessation studies in developed countries, with promising results. Given the exponential growth of mobile phone usage in Vietnam in recent years, mobile health interventions could be a potential strategy to increase smoking cessation in Vietnam. However, substantial cultural adaptations are needed to optimize the effectiveness of these interventions among Vietnamese smokers.

Objective: This study aims to involve qualitative research to inform the development of a mindfulness-based text messaging smoking cessation intervention for Vietnamese smokers.

Methods: A total of 10 focus groups were conducted with 71 Vietnamese male smokers aged between 18 and 65 years (5-9 participants per focus group). Overall, 5 focus groups were conducted with smokers who had the intention to quit (ie, preparation stage of change in the transtheoretical model), and 5 focus groups were conducted with smokers who did not have the intention to quit (contemplation or precontemplation stage). The focus groups were audio recorded, transcribed verbatim, and analyzed using NVivo 12 software (QSR International).

Results: The major themes included smoking triggers, barriers and facilitators for quitting, the perceptions of text messaging and mindfulness approaches for smoking cessation, and suggestions for the development of a text messaging–based smoking cessation program. Common smoking triggers included stress, difficulties concentrating, and fatigue. Frequently encountering other people who were smoking was a common barrier to quitting. However, participants indicated that concerns about the harmful effects of smoking on themselves and their wives and children, and encouragement from family members could motivate them to quit. The participants preferred diverse message content, including information about the consequences of smoking, encouragement to quit, and tips to cope with cravings. They suggested that text messages be clear and concise and use familiar language. Most smokers perceived that mindfulness training could be useful for smoking cessation. However, some suggested that videos or in-person training may also be needed to supplement teaching mindfulness through text messages.

Conclusions: This study provides important insights to inform the development of a text messaging–based smoking cessation program that incorporates mindfulness for Vietnamese male smokers. The results could also be useful for informing similar programs in other low- and middle-income countries.

(JMIR Mhealth Uhealth 2020;8(6):e17337) doi:10.2196/17337)
KEYWORDS
mHealth; mobile health; text messages; smoking; smoking cessation; mobile phone

Introduction
Smoking is the leading cause of preventable deaths worldwide and a major risk factor for most noncommunicable diseases [1,2]. It has been estimated that smoking would cause 150 million deaths globally by 2030, and 80% of these deaths would occur in low- and middle-income countries [3]. Helping smokers to quit smoking is considered by many to be the only practical way to avoid a substantial proportion of tobacco-related deaths worldwide before 2050 [4].

With 15.6 million smokers [5], Vietnam (a lower-middle-income country in Southeast Asia) is one of the top 10 largest cigarette-consuming countries worldwide [6]. Most Vietnamese smokers are men, with a prevalence of smoking of 45.3% among men and 1.1% among women in 2015 [5]. Smoking imposes a tremendous health and economic burden on Vietnam [7,8]. Over 66,000 deaths in Vietnam were attributed to major tobacco-related diseases in 2013 alone [9]. Unfortunately, smoking cessation programs are still scarce [10], even though Vietnam has ratified the World Health Organization Framework Convention on Tobacco Control since 2004. Smoking cessation programs that incorporate evidence-based strategies and are culturally appropriate and cost-effective are urgently needed in Vietnam.

Mobile health (mHealth) interventions have tremendous potential for cost-effective dissemination of smoking cessation interventions. In particular, text messaging programs enable the provision of tailored advice and support, which can be delivered at a relatively low cost, and are scalable for a wide public health impact [11]. Recent systematic reviews show that the quit rates among smokers receiving text messaging–based interventions were consistently higher than those in control groups, with odds ratios ranging from 1.35 to 2.89 [11-14]. However, the majority of these published studies were conducted in high-income countries and none in Vietnam. In 2018, there were 147 mobile phone subscriptions per 100 people in Vietnam, which exceeded that of high-income countries (126 per 100 people) [15]. Moreover, a recent study found that 72% of Vietnamese smokers with the intention to quit were willing to use and pay for smoking cessation support via text messages if available [16]. Thus, text messaging appears to be an acceptable modality for smoking cessation in Vietnam. However, formative research is needed before developing an intervention to ensure that the language, content, and format are culturally appropriate.

Mindfulness training could be an important culturally relevant strategy that has not yet been examined in smoking cessation interventions in Vietnam. Mindfulness is defined as purposeful, present-focused attention, with an attitude of acceptance and nonjudgment [17,18]. Mindfulness originated in Buddhist traditions and is central to the Theravada Buddhist tradition and Mahayana (Zen) schools of Vietnam [19]. Mindfulness-based interventions show promise for promoting smoking cessation in diverse populations in the United States [20]. These interventions teach smokers to notice and pay attention to their emotional states and their cravings with a nonjudgmental attitude. For example, rather than impulsively reacting to cravings by smoking, mindfulness can help smokers to consciously choose how to respond in healthier ways [21]. Mindfulness practices such as mindful breathing (focusing on one’s breath) do not require a high level of education or resources. Therefore, a text messaging program with the integration of mindfulness messages might be beneficial in low-resource settings.

Given that no published research has examined either mindfulness or mHealth for smoking cessation in Vietnam, this qualitative study sought to gain feedback from Vietnamese smokers to inform intervention development. Specifically, this study aimed to understand smoking triggers and barriers and facilitators for quitting smoking among Vietnamese male smokers. We also examined the perceptions of participants about mindfulness and text messaging strategies for smoking cessation, including questions about the content and frequency of text messages. The results of this study will be used to inform the development of a mindfulness-based mHealth smoking cessation program in Vietnam, and the findings might also be useful for guiding similar interventions in other low- and middle-income countries.

Methods

Design
A total of 10 focus groups were conducted among adult Vietnamese male smokers, half (5 focus groups) of which were conducted with smokers who intended to quit in the next 30 days (ie, in the preparation stage of change of the transtheoretical model [22]) and the rest with smokers who did not have intentions to quit in the next 30 days (in the contemplation or precontemplation stage; see details in Participants and Recruitment).

Participants and Recruitment
Participants were contacted by recruiters (health staff at commune health stations) and were asked to complete the screening questions to determine their eligibility and stage of change. Individuals were eligible if they were Vietnamese men between the ages of 18 and 65 years and reported that they currently smoked cigarettes. Female smokers were not included in this study because only 1.1% of Vietnamese women are smokers, compared with 45.3% among men [5]. Smokers who reported intentions to quit within the next 30 days and had at least one 24-hour quit attempt in the past year were classified into the preparation stage [23]. Smokers who were either not interested in quitting or indicated an interest in quitting in the next 6 months (but not the next 30 days) were classified into the precontemplation and contemplation stages, respectively, [23].

Data Collection and Procedures
Data collection occurred between March and April 2018. The quantitative survey questionnaires and the focus group
moderator guide were developed by the research team in English and then translated to Vietnamese. The research team provided detailed information about the study and obtained written informed consent from all participants. Before participating in the group discussion, the participants completed a short survey on their sociodemographic background, smoking habits, cell phone usage (ie, type of cell phone, internet access, and texting habits), and level of nicotine dependence (Fagerström test for nicotine dependence; classified into four levels: low, low to moderate, moderate, and high [24]).

Each group discussion lasted between 90 and 120 minutes. The facilitator followed a semistructured interview guide that asked the participants about their smoking triggers; barriers and facilitators to quitting; level of interest in text messaging to help them quit smoking; suggestions for the content of the messages, including mindfulness-based messages; preferences for the structure and timing of the messages; and other ideas for making the program more helpful and user-friendly. Some technical terms were explained to the participants to ask for their suggestions, such as interactive messages (ie, users would interact with the SMS system by answering questions and receiving more information based on their answer) or keywords (users could send keywords such as “stress”, “crave”, or “slip” to get extra support for coping with stress, craving, or smoking lapses). The concept of mindfulness (translated as mindfulness in Vietnamese) was explained to the participants, and then the participants watched a short video about mindful breathing that led them into a 3-min mindful breathing practice. After the mindfulness practice, participants were asked about their understanding of mindfulness and their thoughts about incorporating mindfulness into smoking cessation treatment. All group discussions were audio recorded using a digital voice recorder. This study was approved by the institutional review board of Georgia State University (reference number H18030).

Data Analysis

Descriptive statistics were applied for quantitative data analyses (ie, frequency and percentage for categorical variables and mean and range for continuous variables). With regard to qualitative data, all recordings were transcribed verbatim and then translated into English for coding. The transcripts were managed and coded using NVivo 12 software (QSR International). Data coding and analysis followed both inductive and deductive approaches [25,26]. A coding manual was first developed by 2 research team members, based on the interview guide and other recurring concepts in the focus group transcripts. The coding manual was refined through group discussions among all research team members. The remaining transcripts were then independently coded by 2 members who were bilingual in Vietnamese and English. During the coding process, regular meetings that included all research team members were held to resolve coding discrepancies, maintain coding consistency over time, and further refine the coding scheme as needed. Discrepancies were resolved through group discussions and consensus among all research team members.

Results

Characteristics of the Participants

A total of 71 adult male smokers participated in 10 focus groups, 5 of which were conducted among 37 smokers with the intention to quit, and the other 5 with 34 smokers who had no intention to quit. The demographic and background characteristics of these 71 participants are presented in Table 1. The mean age of participants was 42.2 years. Approximately 80% of the participants had a cell phone that could access the internet. Half of the participants reported that they often checked new messages immediately as they are delivered. More than 90% of the participants were daily smokers, and they smoked 13 cigarettes on average per day. Most participants had moderate-to-high levels of nicotine dependence based on the Fagerström test for nicotine dependence. Although smokers who were single were 3 times as likely to be in the contemplation or precontemplation stage, there were no substantial differences in sociodemographic or smoking-related characteristics between the participants in the preparation stage and the contemplation or precontemplation stage.
Table 1. Participant characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Preparation stage (n=37)</th>
<th>Contemplation or precontemplation stage (n=34)</th>
<th>Total (N=71)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (range)</td>
<td>42.7 (19-56)</td>
<td>41.7 (18-64)</td>
<td>42.2 (18-64)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than secondary school</td>
<td>2 (5)</td>
<td>2 (6)</td>
<td>4 (6)</td>
</tr>
<tr>
<td>Secondary school completed</td>
<td>9 (24)</td>
<td>9 (27)</td>
<td>18 (25)</td>
</tr>
<tr>
<td>High school completed</td>
<td>23 (62)</td>
<td>19 (56)</td>
<td>42 (60)</td>
</tr>
<tr>
<td>University/college</td>
<td>3 (8)</td>
<td>4 (12)</td>
<td>7 (10)</td>
</tr>
<tr>
<td><strong>Working status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government employee</td>
<td>8 (22)</td>
<td>7 (21)</td>
<td>15 (21)</td>
</tr>
<tr>
<td>Factory, business, or service industry employee</td>
<td>16 (43)</td>
<td>10 (29)</td>
<td>26 (37)</td>
</tr>
<tr>
<td>Farmer</td>
<td>12 (32)</td>
<td>8 (24)</td>
<td>20 (28)</td>
</tr>
<tr>
<td>Student</td>
<td>1 (3)</td>
<td>6 (18)</td>
<td>7 (10)</td>
</tr>
<tr>
<td>Retired</td>
<td>0 (0)</td>
<td>2 (6)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Unemployed, able to work</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>1 (1)</td>
</tr>
<tr>
<td><strong>Marital status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>4 (11)</td>
<td>9 (27)</td>
<td>13 (18)</td>
</tr>
<tr>
<td>Married</td>
<td>33 (89)</td>
<td>25 (74)</td>
<td>58 (82)</td>
</tr>
<tr>
<td><strong>Cell phone can access the internet, n (%)</strong></td>
<td>26 (72)</td>
<td>29 (85)</td>
<td>55 (79)</td>
</tr>
<tr>
<td><strong>Frequency of checking text messages, n (%)</strong></td>
<td>20 (56)</td>
<td>17 (50)</td>
<td>37 (53)</td>
</tr>
<tr>
<td>Immediately as they are delivered</td>
<td>2 (6)</td>
<td>1 (3)</td>
<td>3 (4)</td>
</tr>
<tr>
<td>About every hour during the day</td>
<td>4 (11)</td>
<td>3 (9)</td>
<td>7 (10)</td>
</tr>
<tr>
<td>At least 4 times a day, but not every hour</td>
<td>7 (19)</td>
<td>6 (18)</td>
<td>13 (19)</td>
</tr>
<tr>
<td>2-3 times a day</td>
<td>1 (3)</td>
<td>1 (3)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Once a day</td>
<td>2 (6)</td>
<td>4 (12)</td>
<td>6 (9)</td>
</tr>
<tr>
<td>Never</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Daily cigarette smokers, n (%)</td>
<td>34 (92)</td>
<td>32 (94)</td>
<td>66 (93)</td>
</tr>
<tr>
<td>Number of cigarettes per day, mean (range)</td>
<td>12 (2-25)</td>
<td>15 (1-400)</td>
<td>13 (1-400)</td>
</tr>
<tr>
<td><strong>Nicotine dependence, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>3 (8)</td>
<td>7 (21)</td>
<td>10 (14)</td>
</tr>
<tr>
<td>Low to moderate</td>
<td>11 (30)</td>
<td>5 (15)</td>
<td>16 (23)</td>
</tr>
<tr>
<td>Moderate</td>
<td>6 (16)</td>
<td>6 (18)</td>
<td>12 (17)</td>
</tr>
<tr>
<td>High</td>
<td>17 (46)</td>
<td>16 (47)</td>
<td>33 (47)</td>
</tr>
</tbody>
</table>

Qualitative Results

The major themes were smoking triggers, barriers to quitting and reasons for not wanting to quit, facilitators for quitting, perceived usefulness of text messaging for smoking cessation, suggestions for text message content, frequency and timing of message delivery, text messaging program duration, interactivity of text messages, perceptions about mindfulness and mindfulness-based methods for smoking cessation, and suggestions for incorporating mindfulness into smoking cessation programs.

Smoking Triggers

Smoking triggers were diverse among smokers in both groups. Most participants reported that they often wanted to smoke when they experienced stress or needed to concentrate or fight against sleepiness. In addition, many smokers indicated wanting to smoke when drinking alcohol, tea, or coffee by themselves or with friends at parties or coffee shops. Other common triggers included seeing someone smoking, seeing smoking friends, or being offered cigarettes. Some less common smoking triggers included cold weather and feeling bored:

I usually smoke when I need to concentrate on work, when I’m stressed, when seeing others smoke, it...
reasons for not wanting to quit (among smokers without the intention to quit)

Believing that quitting is very hard or almost impossible was the most common reason for not wanting to quit among smokers who were in the contemplation or precontemplation stages. Many smokers reported that they did not want to quit because they needed to smoke to cope with stress/pressure from work, concentrate, or fight against sleepiness when working during night shifts. Other barriers to quitting included having friends who smoke and the availability of cigarettes:

Being addicted already, there must be a chance or a drug that could decrease the desire, the craving; therefore, giving up consequently. Just quit, how to quit? Quit, you cannot. [46 years, contemplation/precontemplation stage]

I used to have the intention to quit smoking, two times, but haven’t been able to make it successfully because after quitting, I kept returning to the cigarettes when working at night. I also don’t know the ways as to how to quit smoking, how to make me less want to smoke. My aspiration for quitting smoking is still there, but I don’t know how to achieve that, so I keep smoking. [58 years, contemplation/precontemplation stage]

barriers to quitting (among smokers intending to quit)

Among smokers with intentions to quit, a common barrier to quitting was frequently encountering other smokers or smoking friends (eg, seeing someone smoking, smelling smoke, receiving cigarette offers), which made it hard to control their cravings. Others found that the inability to cope with nicotine withdrawal symptoms was a barrier to quitting smoking, including bad or bland taste in their mouth, a feeling of missing something, concerns about gaining weight, or not being able to concentrate. In addition, some participants reported that they needed to smoke when they felt stressed or sad. Some smokers perceived their smoking as a long-time habit that was difficult to give up:

Actually, I want to quit smoking for a long time. It is hard. If I quit, I am not able to hang out with my friends. It will be a hard time because I cannot join my drinking buddies. It is only possible when I am alone or at home. It is impossible to quit if there are people smoking around me. [49 years, preparation stage]

facilitators for quitting

Many smokers in both groups indicated that concerns about the harmful effects of smoking on themselves and their wives, children, or other people around them could motivate them to quit smoking. Complaints and encouragement from family members, especially wives and children, or friends and colleagues were also perceived as facilitators for quitting for many smokers, regardless of their stage of change. Other facilitators included medication to help cope with cravings, being in nonsmoking environments (eg, banning smoking in public places, staying away from other smokers), having a friend to quit smoking together, and being told by a doctor that they have a serious disease as a result of smoking (eg, cancer). Some smokers in both groups indicated that nothing could help them quit smoking, and it all depended on their own will or determination:

My motivation is that I worry about the health problem of myself and people around; I tend to actively participate in social and sports activities to cut down on stress in order to overcome the desire of smoking cigarettes. [33 years, preparation stage]

First, brothers, peers, friends, and neighbour’s encouragement. For example, your children said that you have to stop smoking, this and that - for example, it is also bad for health. Secondly, for instance, there is a drug in order to... just like I said - to help quit smoking - to reduce the craving, if you are determined, it may be combined, then you can quit; if somebody is just thinking about to quit, it will be difficult. Self-quitting is hard. I feel it pretty hard. [46 years, contemplation/precontemplation stage]

level of interest in sms text messaging smoking cessation programs

Overall, the majority of smokers with intentions to quit indicated a high interest in receiving text messages that support smoking cessation. They believed that messages that encourage them to quit or provide them with more information could be helpful. However, many smokers without intentions to quit said that text messaging would not be helpful because they perceived quitting to be impossible or believed that their ability to quit only depended on their determination.

concerns about sms text messaging smoking cessation interventions

When asked about their concerns about text messaging to quit smoking, some smokers thought that there had already been many educational campaigns about smoking (eg, communication campaigns about the harms of smoking via radio, television, newspapers, posters, billboards, or label warnings with scare tactics), and therefore the text messaging program would not be effective if it were similar to these other communication programs. In addition, many smokers worried about receiving too many messages, which could be irritating as they already received too many spam messages. There were also concerns that a text messaging program might not be suitable for older people who might be less familiar with text messaging or have difficulty reading messages of a small font size. Finally, participants who owned basic cell phones with small black-and-white screens reported being less interested in reading messages:
Suggestions for Message Content

When asked about the content of messages that would help to encourage smokers to think about or actually quit smoking, the negative effects of smoking was the most common suggestion of smokers in both groups. Many believed that messages should mention the harmful effects to smokers and their family, and some smokers thought that they should include concrete evidence or specific examples (eg, the number of people who died because of smoking or the number of people who died from cancer caused by smoking, how smoking affects health, diseases caused by smoking). A few smokers suggested that fear-inducing content about the effects of smoking would motivate them to quit:

We should say that smoking affects every aspect of a person’s life, health, and people around them. The wording should make them frightened. It has to involve family, only then they will think that the smoke can affect their wife, children, and the people around them. [53 years, preparation stage]

In the content, we can refer to health warnings, then warnings about the effects smoking might have on people around, their closest people like their parents, wife, and children. My wife also doesn’t like me to smoke, because when I smoke, my breath stinks, and she doesn’t like that. [48 years, contemplation/precontemplation stage]

Smokers in both groups suggested that text messages should provide quitting tips and encouragement, including strategies for coping with cravings. Some participants thought that smokers already knew about the harmful effects or health risks of smoking, and therefore, the messages should not focus too much on that issue but further highlight the effects of smoking on other family members (eg, wife and children). Other less common content suggestions included information about smoking cessation medication, the impact of smoking on finances, and the benefits of quitting to smokers:

I think the content of the messages should refer to how we can reduce the stress, how to deal with the craving for cigarettes when you’re upset and want to smoke, like instructing people what to do in order to achieve the aforementioned goals, taking a bath or meditation, since each person has their own circumstances and it’s not like everyone can sit whenever they want to. [46 years, preparation stage]

Men are the pillars of family. First, we worry about our health, how we can work to earn money to support our family, wife, and children. You can deliver messages about the effects on our health if we couldn’t quit, it’ll affect not only our health but also your family. [29 years, contemplation/precontemplation stage]

Some smokers indicated that including photos in the text messaging program, images showing diseases caused by smoking (similar to graphic warning labels on cigarette packages), could be helpful. Images could also show successful cases of quitting smoking or the happiness of a family with a former smoker. Some smokers suggested that the program could send videos about smokers who are suffering or about the health consequences of smoking:

The content of the images and videos will be similar to the messages you want to deliver. If you want to propagate about the fact smoking can cause lung cancer, you should use images of a lung cancer case to shock people. We can tell how horrible it is after just one glance at that. Or use a video with scenes of a patient trying to get this disease cured or add a report or conversation between a doctor and a patient who has lung cancer because of smoking. I think that way, they will be impacted more directly. [41 years, contemplation/precontemplation stage]

Frequency of Message Delivery

Smokers with intentions to quit smoking wanted to receive messages more frequently than smokers without the intention to quit. Most smokers with intentions to quit said that receiving 1-2 messages per day would be appropriate if the content of the messages is diverse. Some smokers in this group suggested having 3 messages per day. Only one smoker said that one message every 4 days would be enough. Some smokers suggested that the frequency could be decreased over time (eg, 2-3 messages per day in the first month and then decrease to 1 per day or 1 per 2 days in the subsequent months):

One message in the morning for the whole day. If three messages have the same content and purpose, people only need to read one then understand. For example, you can send three messages with different contents. Unless we send three messages with different information, one message is enough. This is my personal opinion. [36 years, preparation stage]

One message per day in the evening, but the content should be diverse. Many messages have such poor content. If it comes during work hours, I’d be mad. [38 years, preparation stage]

In the group of smokers who did not have the intention to quit, the suggested frequency varied from 4 messages per day to 3 messages per month. The most common suggestion was sending messages every day (1 to 3 messages per day), followed by sending 2 to 3 messages per week:
I think the number of messages should decrease gradually by months, one message every two days in the morning at first, then in the second month it will be one to two messages per week. The program should last for about one year and the messages should be sent in the early morning. [29 years, contemplation/precontemplation stage]

Timing for Message Delivery
Most smokers in both groups suggested that they would prefer receiving messages in the early morning (from 6 AM to 8 AM). Evening time (after dinner and before bedtime, 7 PM to 9 PM) was also suggested by many smokers. Some smokers mentioned that they often smoked in the early morning after waking up; therefore, sending messages at that time would help remind and encourage them to stay smoke-free. Sending messages in the evenings about the negative effects of smoking would make them think more about quitting. Other suggestions included sending messages after working hours in the morning and afternoon, when smokers often gather to smoke together, or after lunch, before working hours in the afternoon:

I often smoke in the early morning. So, if you send out messages at six or half past six a.m., and I receive the message while I’m smoking, maybe I can smoke less in the morning. [38 years, preparation stage]

The most effective, I think, the moment is in the morning, and before bedtime, the message notifying me about the effect [of smoking] is usually in the evening. Imagine, when I have a big cough, the message comes, I find it helpful and this, the harmful cigarette, is so true. [48 years, contemplation/precontemplation stage]

Duration of the Text Messaging Program
Among those with intentions to quit, the majority thought that the program should last for about 3-6 months. Many smokers thought that it should last for a year to help them quit smoking. Some smokers thought that it should last for 1 month. Only a few smokers indicated that the program should last for just 1 week to 10 days:

I think we should send one or two messages per day, in about three months. That’s enough time for people to absorb everything. [55 years, preparation stage]

I think there should be two messages a day at an early time. The first period of time is about three months, and we send two a day. After three to six months, we send one message a day or one per two days. After they could understand, then we should also decrease. But in general, the intervention should last for a year. [32 years, preparation stage]

Among smokers who had no intention to quit, most thought that the program should last for a year to help them quit smoking effectively. The second most common suggestion was 3-6 months. The shortest duration was 1 month but was only suggested by a few smokers in this group:

It should last for about one year, in my opinion, it’s up to [individual’s] attitude, it’s no difference even for longer time. [23 years, contemplation/precontemplation stage]

Interactive Messages
Almost all smokers in both groups thought that interactive messages would be helpful for quitting. Only a few smokers who had no intention to quit thought that sending/receiving interactive messages would not be helpful to quit smoking:

That [interactive message] will be more helpful, more helpful. And so, when you are in the stage of craving for a cigarette, but now you wrote a message to send to see how to reduce the craving. [57 years, preparation stage]

As far as I understand, two-way communication is always much better. My point of view is, I strongly agree, for people to have two-way information, there are things that we need advice, if we need to confer directly, then ask immediately. [52 years, contemplation/precontemplation stage]

One smoker suggested that there should be a keyword list for smokers. When asked about suggestions for the keywords to be included on the list, smokers found it hard to give suggestions. However, some smokers suggested that they wanted to be able to request help via text messaging when they had cravings (eg, while drinking alcohol or feeling sad) or nicotine withdrawal symptoms. It was suggested that the program could consider developing a list of keywords based on the list of nicotine withdrawal symptoms in addition to basic keywords such as stress, crave, or slip:

My idea is that the switchboard could send a list, a list of motivated points to get rid of cigarettes. Example A: Stress, B: Tired, C: problems related to the mental/physiological issues in order to give up cigarettes. Then the receiver will be aware of these issues. According to that message, people could request for advice. [33 years, preparation stage]

Perceptions About Mindfulness
After listening to the explanation about the concept of mindfulness and having a short mindful breathing practice, most participants perceived mindfulness as a method that helps to relax or reduce stress or tension and calm the mind, or pay attention or focus on doing one particular thing to forget cravings or other things. Some other smokers related mindfulness as a method to help them to establish a better habit to replace smoking, a method that helps focus on reality and avoid thinking about nonsense or a method that helps people forget their cravings. Some other smokers understood mindfulness as a meditation method or yoga practice:

It is the approach that helps you avoid stress, and relax your mind, with comfortable mind you don’t think much about other things, you are not affected and stimulated. [32 years, preparation stage]

As far as I think, this is to create for you a habit to focus on one thing to forget other things, which means mindfulness, it is now you pay your attention to what is happening and ignore the other things, I mean that

http://mhealth.jmir.org/2020/6/e17337/
is mindfulness. I think it is also good; there is no problem then. [54 years, contemplation/precontemplation stage]

**Mindfulness for Smoking Cessation**

When the participants were asked about including mindfulness as a part of the program, almost all smokers thought that mindfulness would be very helpful for smoking cessation. They thought that it would help them overcome cravings, think less about smoking, or reduce stress, which is a common trigger for smoking:

*This method helps us relax, make our mind comfortable, undisturbed, and not to worry about anything. No irritation, no thought about smoking in our mind, no feelings of craving. I think this is a wonderful method. It’d be extremely good if we can apply this method. Just sitting like that and totally relax, not thinking about anything. I think it’s something like that.* [55 years, preparation stage]

*Before the discussion, everyone usually says that it’s the stress from work, irritation, and sorrow that make them stressed, now that this mindfulness can help reduce their stress and eliminate all those three factors. After erasing them, of course, we can smoke less, I think it’s really good.* [58 years, contemplation/precontemplation stage]

However, a few smokers thought mindfulness was not helpful for smoking cessation or had concerns about applying mindfulness-based approaches. One of these concerns was that practicing mindfulness might be difficult, especially for impatient people or smokers who have been smoking for a long time. In addition, longer formal mindfulness practices that take more time (eg, long bouts of sitting meditation or yoga) might not be feasible for everyone. Some smokers indicated that they would not be interested in mindfulness if it were described as a religious practice. Smokers who perceived the need to smoke as a way to stay alert or cope with fatigue said that mindfulness would be very helpful for smoking cessation. They thought that it would help them overcome cravings, think less about smoking in their mind, and feel more comfortable, undisturbed, and not to worry about anything.

*This method helps us relax, make our mind comfortable, undisturbed, and not to worry about anything. No irritation, no thought about smoking in our mind, no feelings of craving. I think this is a wonderful method. It’d be extremely good if we can apply this method. Just sitting like that and totally relax, not thinking about anything. I think it’s something like that.* [55 years, preparation stage]

*Before the discussion, everyone usually says that it’s the stress from work, irritation, and sorrow that make them stressed, now that this mindfulness can help reduce their stress and eliminate all those three factors. After erasing them, of course, we can smoke less, I think it’s really good.* [58 years, contemplation/precontemplation stage]

**Suggestions for Incorporating Mindfulness Into Smoking Cessation Programs**

Many smokers thought that mindfulness should be incorporated into smoking cessation programs using videos (eg, posted on websites such as YouTube or Facebook) together with some written instructions. Mindfulness could be practiced through simple activities, such as walking, playing sports, or gardening, as formal meditation practice might not be feasible for every smoker. Some smokers were concerned that a program that only uses text messages (without photos or videos) to teach mindfulness would not be sufficient, as each message has a limited number of characters, and mindfulness is a relatively new concept that might be difficult to understand at first. One strategy might be to encourage smokers to practice mindfulness (eg, suggestions for focusing on the present moment and coping with cravings without automatically reacting by smoking), without necessarily using the term mindfulness, which could be confusing. Some smokers suggested that in-person instruction and guidance would be helpful for learning about mindfulness.

**Discussion**

**Principal Findings**

This study examined perceived smoking triggers, barriers and facilitators to quitting smoking, and perceptions about SMS text messaging and mindfulness-based smoking cessation methods among Vietnamese male smokers. Several important findings emerged: first, the majority of Vietnamese smokers with intentions to quit in our study, but only a few of those without the intention to quit, expressed interest in the SMS text messaging program. Second, regardless of their intentions to quit, participants indicated that the message content should be diverse to avoid repetition. They suggested that the messages could include (1) information about the harmful effects of smoking not only on smokers but also on the people around them (eg, wives and children), (2) facts about the consequences of smoking with specific data or evidence, and (3) positive encouraging messages, particularly effective tips for coping with cravings. In addition, each message should be concise, use language familiar to smokers, and avoid ambiguous or confusing words. This is particularly important because standard text messages that are compatible with all types of cell phones in Vietnam do not include accents, and without accents, the meaning of words in the Vietnamese language can sometimes be misunderstood. Third, many smokers preferred that messages be sent every day, with 1 to 2 messages per day, preferably in the early morning and evening. In addition, many smokers suggested that a text messaging–based smoking cessation program should last for approximately 3 months. Finally, many Vietnamese smokers were interested in the application of mindfulness in smoking cessation. Most participants perceived that mindfulness could potentially help with quitting, but some suggested that additional resources may also be needed to supplement teaching mindfulness through text messages.

**Comparison With Prior Work**

The most commonly perceived smoking triggers were consistent with those reported in studies with other populations of adult smokers, including stress [27-30] and drinking alcohol [31]. In addition, beverages such as tea or coffee, popular drinks of Vietnamese people, can be potent triggers for many Vietnamese smokers. Smoking cessation interventions should encourage smokers to think about their own personal triggers, which might include culturally specific triggers (eg, social events where smoking is common among Vietnamese men), and plan ways to avoid and/or cope with them.

Among smokers who had no intention to quit, the most common reason was the perception that quitting is very difficult or impossible. This perception might be a result of their failure in their past attempts to quit or because of the lack of effective methods to deal with cravings and withdrawal symptoms. Other quitting barriers found in both groups of smokers were similar to those reported in previous studies, such as perceiving the need to smoke for stress management, concentration, cravings, and withdrawal symptoms [32]. Smokers noted that an awareness of the harmful effects of smoking was an important motivator for quitting, which is similar to other studies, highlighting that an awareness of the negative effects of smoking

http://mhealth.jmir.org/2020/6/e17337/
might lay the foundation for progress toward smoking cessation [33]. In addition, smokers perceived that complaints and encouragement from other people would also help motivate them to quit. Consequently, smoking cessation interventions may need to focus on increasing the self-efficacy of smokers who have no intention to quit and provide them with strategies to cope with cravings, at the same time, encouraging smokers who have the intention to quit to elicit support from their family members or friends.

With regard to the development of an SMS text messaging smoking cessation program, we found that strategies or tips for managing their cravings and motivational or encouraging messages were the most common desired content among the smokers in our study. This is consistent with the findings from previous studies on adult smokers in the United States [34,35]. In addition, Vietnamese smokers in our study valued content about the harmful effects of smoking not only on smokers themselves but also for their family members. Smokers also indicated a preference for specific evidence, facts, or numbers that could motivate them to quit and remain smoke-free. The inclusion of photos, videos, and interactive functions was valued by most smokers. However, despite the ubiquitous interest in having photos or video links included in the messages, these formats might only be compatible with smartphones. Interestingly, we found that the use of scare tactics suggested by smokers in the study by Bock et al [34] might not work for Vietnamese smokers as the smokers in our study indicated that health communication programs and warning labels on tobacco packs already used such tactics, and therefore additional messaging might not have strong effects. Another difference between our study and previous qualitative studies of text messaging for smoking cessation (most of which were conducted in the United States or other high-income countries) is that most smokers in our study did not want to receive more than 2 messages per day. Studies indicate that smokers in developed countries may prefer to receive more messages, even 5 to 6 messages per day around their quit date [12,34,35]. This difference might be explained by the large number of spam messages sent to cell phone subscribers in Vietnam, which made Vietnamese cell phone users reluctant to receive text messages. The 3-month duration of the text messaging program preferred by smokers in our study is consistent with that reported by Bock et al [34] and aligned with implemented trials [12].

After trying a brief mindful breathing practice, participants described mindfulness as a method that could help them relax, reduce stress or tension, calm their mind, or improve concentration. Overall, participants had a positive attitude toward mindfulness, and they believed that this method would be helpful to quit smoking. However, it may be important to clarify with this population that although mindfulness can help reduce stress and anxiety [36,37], it is not the same thing as relaxation. In fact, given that mindfulness involves nonjudgmental attention to events occurring in the present moment, practicing mindfulness can be very uncomfortable (eg, when noticing sensations of craving, pain, or unpleasant emotions). In addition, participants noted that text messages might not be sufficient for teaching mindfulness, given that the concept of mindfulness was relatively new to most Vietnamese smokers in this study. Moreover, even in intensive in-person training programs, participants typically do not practice mindfulness on their own as much as instructed [38]. Text messages could be a useful modality for reminding people to practice mindfulness regularly, but it might be optimal to combine text messaging with web-based, telephone-based, or in-person training. For example, text messages have recently been combined with in-person mindfulness-based treatment for smoking cessation in the United States. It might also be helpful to describe mindfulness practices in clear, understandable ways (eg, slowing down to pay attention, focusing on your breathing), without using the term mindfulness, if that causes confusion for some people.

**Limitations and Conclusions**

This study has several limitations. Our study included only Vietnamese male smokers; therefore, the results may not generalize to smoking cessation programs aimed at female smokers or those in other geographic areas. In addition, this study examined the ideas smokers have about receiving text messages for smoking cessation rather than their actual experiences of receiving sample text messages. An important future direction will be to collect qualitative data after Vietnamese smokers have the opportunity to receive a text messaging intervention. However, this study is strengthened by a relatively large sample size for qualitative research; investigation of themes separately by smoking stage of change; examination of mindfulness as a relatively novel potential strategy for smoking cessation in Vietnam, and inclusion of smokers from diverse backgrounds based on age, education level, working status, and place of residence.

Our findings provide important insights into smoking cessation programs for male smokers in Vietnam, a country with a high prevalence of smoking and related morbidity and mortality among men, but low accessibility to smoking cessation treatment. Overall, participants (especially those with intentions to quit smoking) were interested in SMS text messaging as a method to help them quit smoking. They suggested that the messages should be nonrepetitive and have diverse content, including concrete evidence and statistics about the negative effects of smoking for both smokers and their family members. In terms of message frequency, delivering 2 messages per day either during early morning or evening was acceptable to most participants, although an individually tailored delivery of messages catering to specific individual needs would be even better. Mindfulness training was viewed favorably by Vietnamese smokers as a potential solution to help them quit; however, some suggested that videos and/or in-person training may also be needed to supplement teaching mindfulness through text messages.
Acknowledgments

The authors acknowledge the funding support from the Fogarty International Center of the National Institutes of Health under award number R01TW010666. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References


Describing the Process and Tools Adopted to Cocreate a Smartphone App for Obesity Prevention in Childhood: Mixed Method Study

Paolo Giorgi Rossi, PhD; Francesca Ferrari, MSc; Sergio Amarri, MD; Andrea Bassi, PhD; Laura Bonvicini, MSc; Luca Dall’Aglio, MEng; Claudia Della Giustina, MSc; Alessandra Fabbri, MPH, MD; Anna Maria Ferrari, MD; Elena Ferrari, MD; Marta Fontana, MPhil; Marco Foracchia, PhD; Teresa Gallelli, BEd; Giulia Ganugi, PhD; Barbara Ilari, MD; Sara Lo Scocco, PhD; Gianluca Maestri, PhD; Veronica Moretti, PhD; Costantino Panza, MD; Mirco Pinotti, MD; Riccardo Prandini, PhD; Simone Storani, MD; Maria Elisabeth Street, MD, PhD; Marco Tamelli, PhD; Hayley Trowbridge, PhD; Francesco Venturelli, MD; Alessandro Volta, MD; Anna Maria Davoli, MD; Childhood Obesity Prevention Working Group

1Servizio Epidemiologia, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
2Pediatria Arcispedale Santa Maria Nuova, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
3Dipartimento di scienze politiche e sociali, Università di Bologna, Bologna, Italy
4Servizio Tecnologie Informatiche e Telematiche Interaziendale, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
5Servizio Igiene degli Alimenti e Nutrizione, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
6Dipartimento di Salute Pubblica, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
7Pediatra di libera scelta, Dipartimento di cure primarie, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
8Lepida ScpA, Bologna, Italy
9Medicina dello Sport e Prevenzione Cardiovascolare, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
10Dipartimento di cure primarie, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
11Servizio di igiene pubblica, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy
12People’s Voice Media, Manchester, United Kingdom
13Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy
14Cocreation of Service Innovation in Europe Project, Reggio Emilia, Italy

Corresponding Author:
Francesca Ferrari, MSc
Servizio Epidemiologia
Azienda USL-IRCCS di Reggio Emilia
via Amendola 2
Reggio Emilia
Italy
Phone: 39 0522 335470
Email: francy.ferr80@gmail.com

Abstract

Background: Childhood obesity prevention is a public health priority in industrialized countries. The Reggio Emilia Local Health Authority has implemented a program involving primary and secondary prevention as well as the care of obese children. There are many health-promoting mobile apps, but few are targeted to children and very few are sponsored by public health agencies.

Objective: The goal of the research was to describe the process and tools adopted to cocreate a mobile app sponsored by the Reggio Emilia Local Health Authority to be installed in parents’ phones aimed at promoting child health and preventing obesity.

Methods: After stakeholder mapping, a consulting committee including relevant actors, stakeholders, and users was formed. Key persons for childhood obesity prevention were interviewed, focus groups with parents and pediatricians were conducted, and community reporting storytelling was collected. The results of these activities were presented to the consulting committee in order to define the functionalities and contents of the mobile app.
Results: Three key trends emerged from community reporting: being active, playing, and being outdoors; time for oneself, family, and friends; and the pressures of life and work and not having time to be active and socialize. In focus groups, interviews, and labs, mothers showed a positive attitude toward using an app to manage their children's weight, while pediatricians expressed concerns that the app could increase their workload. When these findings were explored by the consulting committee, four key themes were extracted: strong relationships with peers, family members, and the community; access to safe outdoor spaces; children’s need for age-appropriate independence; and professional support should be nonjudgmental and stigma-free. It should be a dialogue that promotes family autonomy. The app functions related to these needs include the following: (1) newsletter with anticipatory guidance, recipes, and vaccination and well-child visit reminders; (2) regional map indicating where physical activity can be done; (3) information on how to manage emergencies (eg, falls, burns, fever); (4) module for reinforcing the counseling intervention conducted by pediatricians for overweight children; and (5) a function to build a balanced daily diet.

Conclusions: The pilot study we conducted showed that cocreation in health promotion is feasible, with the consulting committee being the key co-governance and cocreation tool. The involvement of stakeholders in this committee made it possible to expand the number of persons and institutions actively contributing to the project.

doi:10.2196/16165

KEYWORDS

childhood obesity; health promotion; mHealth; cocreation; mobile app

Introduction

Obesity, its metabolic consequences (eg, hyperglycemia, hypercholesterolemia), and its risk factors (ie, incorrect diet and low physical activity) are responsible for the vast majority of disability-adjusted life years lost worldwide and in industrialized countries in particular [1,2]. In Italy, about 30% of 8-year-old children are overweight or obese [3]. Childhood obesity is one of the major risk factors for adult obesity and diabetes but also has consequences for the child’s health and well-being [4]. This makes childhood obesity prevention a strategic priority in public health, with potential high impact in the medium and long term. In 2010, the Reggio Emilia Local Health Authority (LHA) started a program of research and interventions aimed at preventing childhood obesity. The program adopted a multilevel and multisetting strategy for primary prevention in the community (particularly in infant-toddler centers and preschools and in primary and secondary schools); secondary prevention (with individual screening for overweight and obese children at age 5 years and counseling by family pediatricians [5,6]; and management of obese children by multidisciplinary teams with treatment of those with complicated pathological obesity. The implementation of such a program increased the need for communication between institutions (LHA, schools, and municipalities), family pediatricians, and families to exchange information about initiatives in school cafeterias, promote physical activity initiatives, facilitate pediatric counseling, and manage multidisciplinary team activities. New information technologies offer the opportunity to open a bidirectional communication channel between parents and institutions to address these needs.

Recently, several health promotion apps have been developed [7-11], which mostly target adults and adolescents [8,10,12]; very few have been produced by governmental institutions [7-11]. Evaluating the efficacy of these apps as public health interventions is challenging as they often include different functions, making it difficult to separate what works from what does not. Furthermore, although different apps share some of the same or similar functions, the apps are not substantially equivalent to each other in terms of their fundamental components. Thus, systematic reviews cannot pool results from different studies [7]. Some studies are available on apps targeting childhood obesity [13,14], but synthesizing and generalizing their results is difficult.

The efficacy of any health promotion effort depends on its ability to reach and engage the target population. New information technology (IT) tools can help by tailoring the intervention and framing the message according to each family’s needs [15,16]. However, in health promotion and preventive care, the beneficiary’s needs and the health service’s aim often do not correspond. This mismatch may be due to the general public’s lack of awareness of the real impact of diet and physical activity on health but also due to the different value that each individual gives to remaining healthy and changing behaviors. There is, therefore, not only the problem of unmet needs but also of unperceived needs and the fact that a health service and the beneficiary may place different values on prevention.

Cocreation is a process to plan and define public services specifically aimed at reducing the mismatch between beneficiary needs and provided services. Its application in health services and prevention has been recommended [17,18], in particular when IT tools are proposed [19-22]. There are published case studies on the development of apps and other eHealth tools [19,23-30].

The aim of this paper is to describe the process and tools adopted for cocreating an app to be installed on parents’ mobile phones aimed at promoting childhood health and preventing obesity.

Methods

Reggio Emilia Pilot Project

This pilot study, using mixed-method research, is one of 9 included in the CoSIE project (Cocreation of Service Innovation in Europe, Call: H2020-SC6-CO-CREATION-2016-2017). This innovative action is aimed at producing guidance on how information and communications technologies can support the process of service cocreation in Europe.
Setting

The Province of Reggio Emilia, located in northern Italy, has a resident population of about 530,000 inhabitants, of whom about 80,000 (15.4%) are children ages 0 to 14 years [31]. The province has 6 health districts, one research hospital and 5 district hospitals, and approximately 90 family pediatricians.

Bambini Molto in Forma Project

The BMInforma project (Italian: Bambini Molto in forma; English: very fit children) is an ongoing multilevel public health program conducted by the LHA involving primary and secondary childhood obesity prevention interventions. It includes all the LHA’s primary prevention routine activities on childhood health promotion, such as development of school cafeteria menus; extracurricular interventions in infant-toddler centers, preschools, and primary and secondary schools in collaboration with municipal educational services; and collaboration with sports associations to promote organized and nonorganized physical activity. Secondary prevention consists of population-based overweight and obesity screening of children aged 5 years. According to the results of a trial conducted locally [5,6], families of overweight girls are invited to participate in a motivational interview program led by the family pediatrician, while overweight boys receive recommendations and body mass index (BMI) monitoring. Obese children are referred to a multidisciplinary team that organizes group interventions involving family pediatricians, dieticians, and psychologists. Obese children with pathological conditions are referred to the pediatric endocrinology unit at the hospital for care of specific pathologies.

This network of services, initially developed in 2011, is still being fine-tuned. Several research projects are nested in this program, such as a cohort study on distal and proximal determinants of childhood obesity and trials to test the efficacy of individual and group interventions for overweight and obese children. Collaboration with other municipal agencies outside the health sector, such as schools, transportation, and city planning, and with nonprofit organizations is becoming more and more important, in line with the indications of the 2014-2019 National Prevention Plan [32].

This background makes the BMInforma project a perfect setting to test innovative and traditional tools for cocreating services with families and for co-governance involving the nonprofit and private sectors and the many province-wide municipal administrations (Figure 1).

Figure 1. Scheme of the interactions between the primary and secondary prevention and obesity care. The figure also depicts the initial project of an app to improve the service network.

Scope of the Needs Assessment Phase

Here we report the results of the needs assessment phase in which we tried to answer the following questions:

- Is the network of initiatives and services on childhood obesity prevention and care meeting the needs of parents and children?
- Are all the components of the network connected and do they share the same objectives?
- How can we improve the network?
- Can an app improve the network?
- What should an app do to be effective?
Cocreation Tools

Stakeholder Map

Stakeholder mapping was conducted using an iterative method. Initially, a restricted group of pilot project coordinators, those involved in drafting the application to Horizon 2020, drafted a first list of internal and external stakeholders, decision makers, and beneficiaries of the pilot project. A template was adopted that included the potential influence/contribution of each stakeholder, potential impact of the project on the stakeholder, and possible strategy to involve each stakeholder. Based on this list, a meeting of all internal stakeholders was organized and a new stakeholder analysis was conducted. This step led to the formation of the project steering committee, which included all main internal stakeholders and two experts as external advisors. Finally, the list of stakeholders obtained in the second step was used to create the consulting committee, which included all external and internal stakeholders, decision makers, and parents. During the first meeting of the consulting committee, a third stakeholder analysis was done to identify other public administration sectors and nonprofit associations conducting related projects. The consulting committee decided to remain open to new participants (ie, stakeholders and/or institutions) for the duration of the project and remain active after the end of the project to coordinate local policies on childhood well-being. The project management design, with the main actors and process of evaluation and feedback, is shown in Figure 2.

Figure 2. Project management design.

Interviews With Key Actors

The aim of the semistructured interviews was twofold: the first was to evaluate the BMInforma project and changes that the project led and the second focused on a new phase of the project, begun with the Horizon 2020 grant, to determine what had already been done and what was planned on the agenda.

The interviewees, 5 family pediatricians and 3 health care professionals, were identified by the steering committee with the support of the Reggio Emilia LHA. In addition, the 3 project managers of an ongoing trial testing the efficacy of educational group therapy for obese children (Gruppi di Educazione Terapeutica) were interviewed to explore new ways to involve participants (children and their family members) in the pilot research project.

The 8 semistructured interviews were carried out by two researchers from the University of Bologna, Italy, between June and July 2018. Each interview lasted about 1 hour; the same question template was used (see Multimedia Appendix 1 for a detailed outline of the interview).

The main themes to emerge during the interviews:

• Role of family pediatricians and health care professionals in the conduction of both projects (BMInforma and CoSIE)
• Expectations of pediatricians and health care professionals regarding the effectiveness of the app
• The value of cocreation

The semistructured interviews were recorded and transcribed verbatim, and interviewees’ sensitive data were anonymized. Qualitative data analysis was performed using NVivo 12 software (QSR International). A content analysis was then conducted to identify the context where specific nuclei of meaning had been expressed. Codes of meaning, links between all the participants’ statements, conceptual frameworks, and interpretative hypotheses were created using the NVivo software.
Focus Groups

Two focus group studies were conducted: one with family pediatricians, the other with parents. The focus group technique, widely used within the social sciences, is an unstructured group interview method that responds to precise rules of preparation, organization, and management. A total of 3 focus group discussions were conducted by the research team from the University of Bologna: one with family pediatricians and two with parents. The mediation of the Reggio Emilia LHA was fundamental in recruiting the participants (both pediatricians and parents). The first focus group discussion was held in May 2018 and involved 14 family pediatricians working in the Province of Reggio Emilia. Thirteen of the pediatricians were female and 1 was male; ages ranged from 30 to 60 years. The focus group discussion lasted about 2 hours (see Multimedia Appendix 2 for a detailed outline of the interview), and the main topics were the role of family pediatricians in dealing with childhood obesity, pediatricians’ relationships with families, and mobile app functions proposed by the consulting committee.

The focus group sessions with the parents (all mothers) were held in October 2018. There were 5 participants in the first and 5 in the second, and ages ranged from 30 to 50 years. Each focus group discussion lasted about 2 hours and the following topics were discussed: lifestyle (how their typical day is organized), nutrition and physical activities, and the role of IT in their lives.

What emerged from the focus groups was analyzed using the same process described for the semistructured interviews, with the addition that word clouds were created through NVivo and used as input during the meetings of the consulting committee.

Public Cocreation Lab

During a national festival on digital innovation that took place in Reggio Emilia on October 20, 2018 (during the needs assessment phase of the pilot study), we organized a cocreation laboratory involving families and professionals called “What do you need on your smartphone for your child’s health?” In a family-friendly environment (with organized entertainment for the children), parents had the opportunity to sit down at any one of 4 tables—on diet, physical activity, communication with family pediatricians, or the relationship with municipal institutions—and talk with experts and decision makers about that topic. The contents of conversations with parents were then summarized by the participant experts. There were also other ways to provide input: four signage totems indicating each of the 4 topics were placed around the space, and parents were invited to leave messages, insights, and comments on Post-it Notes. Also, two tablets were placed in the quietest corners of the room for anyone wanting to leave a video and/or audio message. Last, there were whiteboards available to the children and adults for drawing.

To support family participation, Pause and the Reggio Children’s Foundation organized an atelier dedicated to food and tastes in which children could explore vegetables with all 5 senses. All Post-it Notes, notes from the topic table conversations, and videos were then given to the consulting committee without any pre-analysis.

Community Reporting

Overview

Community Reporting for Storytelling, a pan-European movement established in 2007 by People’s Voice Media, uses digital tools to gather, curate, and mobilize lived experience data. Its methodological approach is based on the Cynefin decision-making framework for complex environments [33].

Adopting the gathering, curating, and mobilizing community reporting cycle, bespoke interventions across 3 stages were designed and implemented within this study to better understand the needs of families in terms of what keeps them well.

Stage 1: Community Reporter Training

Community reporting has three interlinked storytelling models: storytelling, coproduction, and insight. Within this study, the insight approach was used as it provides rich qualitative data to projects by taking the insights from people’s stories to identify a core set of research findings that can be used to inform policy, practice, and service design. Nine participants, including pediatricians, researchers, and members of the pilot project core team, were trained in this approach as part of a 2-day program held May 8-9, 2018, underpinned by peer and experiential learning strategies; the participants explored the following topics:

- The community reporter movement
- Insight storytelling techniques:
  - Snapshot stories: short responses to an open question
  - Dialogue interviews: unstructured and unscripted interviews with only one preset question used as a conversation starter
- Responsible storytelling and cocreated best practice guide
- Sharing stories online

Using these skills, participants videorecorded a set of stories from families (parents and children) about what keeps them well.

Stage 2: Story Gathering and Curation

The community reporters trained in stage 1 gathered more insight stories and uploaded them to the Institute of Community Reporters website [34] from June to September, 2018. The 17 stories gathered during stages 1 and 2 were subsequently analyzed using the Institute of Community Reporters’ analysis model, which examines each story in terms of topic, content, and contextual levels before inductively determining the findings across the stories. In essence, the approach is broadly based on principles associated with established methodologies within discourse analysis [35] and on grounded theory [36].

Stage 3: Mobilizing the Insights in the Stories

A conversation of change activity was run October 30, 2018, using findings and extracts from the stories gathered during stages 1 and 2 as part of a consulting committee workshop. Adopting facilitation techniques that drew on principles of open space technologies, Brené Brown’s vulnerability research and story dialogue techniques [37], stories and findings were used as stimuli for cocreative conversation about what keeps families well.
Analysis and Synthesis

The consulting committee members received all materials collected in the interviews, focus group sessions, community reporting, and cocreation lab during a plenary session workshop. The materials were organized into the main topics by the social science researchers of University of Bologna and by curators of People’s Voice Media.

The first section of the workshop was a conversation of change activity and involved the findings from the community reporter stories and key exemplary extracts used to prompt thinking about family well-being. Learning from the stimuli was grouped into three categories:

- Key messages of the stories
- Key learning from the stories for health care services
- Experiences of the consulting committee members and how the stories relate

From this, a set of unstructured ideas for the mobile app was produced. This learning was taken forward into the second section of the workshop and combined with other inputs, including focus group sessions, interviews, and the public cocreation lab. Using these, the consulting committee worked in subgroups to summarize and develop the materials in three phases: identifying all possible topics related to family well-being, grouping topics into macro areas that should be covered by the app, and transforming needs into contents or functions the app should include.

After the group completed a list of objects/functions the app should have, back-office work was done to produce a list of the app’s requirements to help the technicians of the regional health authority’s IT service produce technical specifications. The subgroups were organized in order to better deal with topics that had similar technical issues. Interventions and services that can be provided to families should be evidence-based and recommended by regional and international guidelines. The choice of interventions to be delivered among those that are evidence-based should be made through a needs assessment phase and an evaluation of sustainability and acceptability in the local context. In this phase, the codesign tools are focus groups, interviews with key actors, and community reporting, all of which are discussed and analyzed by a consulting committee made up of the actors, stakeholders, and users. Synthesis of the input from the needs assessment phase is then conducted by the consulting committee. As soon as different prototypes of the app are delivered, according to the established cocreation approach, interviews with convenience samples of potential target families will be conducted to give feedback and continue the cocreation process. Finally, the app public release will automatically collect feedback from users’ interactions and comments, leading to new iterations of the app (Figure 3).

Figure 3. Cocreation strategies for the design, production, and governance of the app in the Reggio Emilia pilot project on childhood obesity prevention.
Results

Stakeholder Analysis and Definitions of the Steering and Consulting Committees

The final list of stakeholders included 17 organizations or groups within organizations (Multimedia Appendix 3). Most were in the public sector, but there were also many nonprofit groups and associations and some from the private sector, the latter are mainly food industries and food distribution companies. We decided to involve them through their professional/business associations and not individually, with the exception of one company that currently provides meals to the public schools in Reggio Emilia. Parents participated only through the inclusion of their class representative in the school parent councils. We did not find any parent associations that focused on healthy lifestyle, obesity prevention, or child well-being; the associations we did find focused on abuse prevention, issues related to divorced parents, and the protection of minors. This lack of parent association representative made it extremely important to activate other means of cocreation to receive the users’ and beneficiaries’ inputs.

The process of stakeholder mapping was explicitly designed to increase the engagement of stakeholders who were not initially involved in the project. In fact, starting as the object of the mapping, they became active subjects who defined the map itself throughout the different steps (Figure 4). The process began with the steering committee and the BMInforma project, which was initiated by a group within the LHA. Other departments of the local and regional public health sector were subsequently involved. The establishment of the consulting committee allowed the introduction of other sectors important to children’s well-being, including education, municipal administrations, social innovation, transport, sports associations, the food production and distribution industries, and parent representatives. Furthermore, to increase engagement of the wider community included in the consulting committee, the functions of the steering committee were reduced to preparing the consulting committee meetings, while some of the back-office work (eg, summarizing community reports and the findings of focus groups), initially thought to be the responsibility of the steering committee, was conducted by consulting committee subgroups (Figure 2). The cocreation and co-governance activities led to active involvement in project management by all key actors in the other public, private, and nonprofit sectors on the consulting committee, transforming stakeholders in actors.

Figure 4. Evolution of the actors and stakeholders and user map.

Defining the App Contents: Interviews, Focus Groups, Community Reports, and Cocreation Lab

Interviews provided insight for the evolution of the BMInforma project and the cocreation of the app. All health professionals confirmed the project gave them new skills and more awareness of the problem and sharpened existing skills such as providing appropriate treatments. Observing what emerged from the cocreation of the app, interviewees demonstrated their interest in profiling the app user in order to allow each family to directly assess part of the information in their electronic health file. The personal page might include information such as the school menu and vaccination appointments.

The topics that emerged from the parent and pediatrician focus groups had more to do with the use of IT, and particularly of apps, during their daily activities. Mothers showed great readiness to use an app to manage their child’s weight. In particular, the mothers viewed the possibility of creating new recipes (thus stimulating creativity in the kitchen), checking the family’s diet, and learning about opportunities to improve their habits in a positive manner. The pediatricians raised concerns that the app could increase their workload rather than be an
instrument to facilitate work and warned that the app cannot substitute the pediatrician in dealing with the patient’s need.

During the cocreation lab, about 100 Post-it Notes, conversation notes, and 10 drawings were collected. The notes were grouped according to the four topic areas of the lab (diet, physical activity, relationships with public institutions, and communication with pediatricians) to be further analyzed by the consulting committee subgroups. No video or audio messages were collected.

Analysis of the 17 community reporter stories gathered brought to light the following key trends:

- Being active in a variety of ways, often involving play and being outdoors, was seen as important in the families’ lives
- Time to oneself, with family, and with friends was important to supporting the families’ overall well-being
- The pressures of life and work and not having time be active and socialize were detrimental to families’ overall well-being

Two key anomalies were seen. One story identified how volunteering can support well-being, and another about a young girl’s use of a step-counting bracelet revealed how there can be unintended negative consequences of health (and technological) interventions.

When findings were explored during the conversation of change activity, learning across four key themes was extracted by the consulting committee:

- Strong relationships with peers, family members, and the community support well-being
- Access to safe (outdoor) spaces that can be used for unstructured activities (ie, free play) is important to families
- Children need to have age-appropriate independence to realize and actualize their sense of self and enable them to support their own well-being
- Professional support based on discussion and exchange should be nonjudgmental and stigma-free and promote family independence and their autonomy in making decisions about their lives

Translating Needs Into App Content

The consulting committee’s work produced four lists of objects and functions that should be included in the app, divided into main topics: diet and healthy menus, physical activity, relationships with public institutions, and communications with family pediatrician. Several topics were included in more than one list, highlighting the overlap between themes. Small working groups of the consulting committee, integrated with external experts, conducted analyses of the overlapping areas, made observations of technical issues to be addressed, and produced a final list of technical specifications and requirements of the app (Figure 5).

Figure 5. Translating needs into app content.
Table 1 compares evidence-based strategies on weight management identified by Rivera et al [38] in a scoping review on mobile apps for weight management with features found in the CoSIE app, which employs 8 of the 11 strategies defined in the review. Only 2 apps (0.5%) in the scoping review included 8 strategies, with an average having between 1 and 2.

Table 1. Evidence-based strategies, health care expert involvement, and scientific testing in apps for weight management and in the CoSIE app. Adapted from Rivera et al [38] (n=393).

<table>
<thead>
<tr>
<th>Evidence-based strategies</th>
<th>Frequency across included apps</th>
<th>CoSIE app features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-monitoring</td>
<td>35.4</td>
<td>Present: Yes, Comment: Only for weight</td>
</tr>
<tr>
<td>Automatic self-monitoring</td>
<td>10.2</td>
<td>Present: No, Comment: Automatic physical activity self-monitoring was not included to avoid causing the children any stress</td>
</tr>
<tr>
<td>Goal setting</td>
<td>21.4</td>
<td>Present: Yes, Comment: Only for selected users (families with children in obesity prevention program)</td>
</tr>
<tr>
<td>Physical activity support</td>
<td>27.5</td>
<td>Present: Yes, Comment: In terms of availability of organized and nonorganized physical activities in the community (using a map with activities geolocalized)</td>
</tr>
<tr>
<td>Healthy eating support</td>
<td>23.2</td>
<td>Present: Yes, Comment: Through recommended recipes in order to have a balanced diet, and information on seasonality of fruits and vegetables and on their nutrients</td>
</tr>
<tr>
<td>Weight/health assessment</td>
<td>25.4</td>
<td>Present: Yes, Comment: Assessment of body mass index only</td>
</tr>
<tr>
<td>Personalized feedback</td>
<td>1.9</td>
<td>Present: No, Comment: We exclude personalized feedback for all users because this would involve an unsustainable workload for public health care professionals. However, for children in the obesity prevention program, the app will inform the pediatrician of body mass index monitoring, eating, and physical activity habits</td>
</tr>
<tr>
<td>Motivational strategies</td>
<td>7.1</td>
<td>Present: Yes, Comment: Through gamification, news, and specific tools for families with children in obesity prevention program</td>
</tr>
<tr>
<td>Social support</td>
<td>5.3</td>
<td>Present: No, Comment: We exclude online communication with other users since it would involve supervision and control, which we cannot guarantee at the moment</td>
</tr>
<tr>
<td>Health care expert involvement in development</td>
<td>0.3</td>
<td>Present: Yes, Comment: —</td>
</tr>
<tr>
<td>Scientific test</td>
<td>0.8</td>
<td>Present: Yes, Comment: In progress</td>
</tr>
</tbody>
</table>

aCoSIE: Cocreation of Service Innovation in Europe.

Discussion

Principal Findings

The cocreation process described here succeeded in producing a list of the content of the childhood obesity prevention app; the content items proposed by the different cocreation process participants and from the users’ (parents) and beneficiaries’ (children) suggestions were largely consistent. Mixing several cocreation tools, some of which were more conventional (eg, forming a consulting committee) while others were more innovative (eg, community reporting and public laboratories), made it possible to balance input from citizens and institutions.

Finally, the process succeeded in transforming the consulting committee into an active community that brought together different sectors, favoring synergy and operating in a “health in all policies” perspective [39].

Strengths and Limitations

This study is merely descriptive; we cannot rule out that similar or even better results and a similar level of decision sharing could be reached with another process. Furthermore, we have reported the process used to determine app content; we do not yet have any information about how the app will actually be used and whether it will be effective in promoting healthy behaviors in children.

A critical point was parent involvement in the consulting committee. The core group proposed not to include specific parent associations because of their focus on some aspects of parenting; we initially included only a representative of the school parent councils, a choice supported by the other stakeholders on the committee. The limited representation of final users on the consulting committee resulted in an orientation more toward co-governance than cocreation. In the literature, in fact, co-governance does not involve any of the beneficiaries; co-governance refers to the joint participation only of public agencies, the private sector, and nonprofit organizations in decision making and the planning of public services [40-42]. The parents’ suggestions had to be collected with more unconventional cocreation tools (ie, focus groups, community reporting, and public lab), as commonly described in similar experiences [17,43-47]. Nevertheless, as the target of our initiative on obesity prevention is the families themselves, many
of the stakeholders on the consulting committee are parents or grandparents of children in the target age (from newborn to preadolescence) and thus also potential end users, and many of them have also faced the issue of overweight/obesity in childhood, given that the prevalence is close to 30%.

The fact that many members of the consulting committee were simultaneously stakeholders and potential users made our cocreation process different from all the previous experiences of cocreation in health interventions (including apps and eHealth products), where end users provided input on the needs and tested the prototypes but were not involved in the identification of the aims or in the conduction of the project [43-47]. In our pilot project, the process of cocreation through the consulting committee activities gradually eliminated the borders between the core pilot project leader group and the other stakeholders and end users involved. This process became clearer when the tasks originally assigned to the steering committee (ie, synthesizing the needs assessment phase into a document suitable for technical requirements and content of the app) were transferred to working groups made up of consulting committee participants.

Topics That Emerged During the Needs Assessment Phase

We obtained a detailed list of content and requirements suitable for technical development of the app. Adopting strategies proposed in a previous review on weight management apps, the CoSIE app as outlined by the cocreation process fulfills most of the quality requirements (Table 1) [38].

Surprisingly, content and topics families proposed were consistent with those identified by the institutions, with a few exceptions. For example, during the public lab, a couple of parents asked for an online pediatrician to provide quick answers to questions posted through the app; the health care professionals, however, considered this unfeasible. In the scientific literature, we found only one study in which parents could directly contact a dietician or psychologist through the app [48,49]. Although the authors did not report any issues regarding health practitioner workload induced by the app-mediated contacts, the study was conducted on a small sample and scaling up to the whole target population of this intervention was not proposed because it was not effective in reducing fat mass index and changing behaviors [49]. Further, LHA officials expected a request to simplify bureaucracy through the app, but no beneficiary made this request. The main theme emerging from parent input was the need for time and for safe public spaces. This was the main concern of most of the institutions as well, who look for ways to increase physical activity. The shared emphasis was on how to facilitate access to playgrounds and other community spaces for children that are suitable for safe, unstructured play. The app’s map of opportunities, which lists all the places in a given neighborhood where it is possible to play and be physically active, could respond to this request; such a function was not common in other similar apps [50]. Lack of time was a key issue in parents’ requests regarding the promotion of a healthy diet and a positive relationship with food. In this case, the only tool an app can offer is the cookbook, with two ways to access it: according to the ingredients in your refrigerator or to your preferences. Some apps evaluated in scientific reports [15,16,48,49] allow self-monitoring of food consumption and weight status, with automatic feedback on correct energy uptake. While providing feedback on daily or weekly diet is still under consideration in the CoSIE app within the gamification and rewarding function (although not considered a high priority by the consulting committee), any direct contact with a health care professional has been expressly ruled out. In general, stakeholders proposed avoiding any prescriptive approach. The CoSIE app does not propose any diets, physical activity programs, or tutorials, unlike similar apps [15], because behavioral changes should come about thanks to a favorable environment and attitude.

What to do in case of an emergency was another topic both parents and health practitioners requested. This topic is not strictly related to obesity prevention and, to our knowledge, not present in other similar apps, but it was recognized as a way to make the app useful and appreciated by parents. A proposal for gamification of the app emerged particularly from the institutions, who saw it as a way to engage families. Both users and institutions agreed that the app should not be used directly by the child but by the parent only (ie, it is on the parent’s smartphone). This is the main barrier to gamification but also guarantees that any game is played with a parent or at least under the parent’s supervision. The solution was to develop games only for designing a healthy menu and planning the right physical activity over the course of a week.

Cocreation in Health Services: Insights From the Pilot Study

In general, health services have several specific characteristics that make cocreation particularly challenging. Analyses have both highlighted the theoretical benefits of involvement strategies to health care (eg, promoting equity and improvement) and identified numerous tensions and contradictions that play out in practice [17,23,51-54]. Nevertheless, within this project, particular emphasis was placed on improving each phase [55] of the process by harnessing the experiences of experts, citizens, and patients. Here we described the first two phases, co-commissioning and co-design, but we have already planned the next steps of co-delivery and co-assessment of the entire obesity prevention program, not just of the app.

The main difference between public health services and other public services is that the former demand scientific evidence of the efficacy of any intervention before it can be provided. In the case of primary prevention, we often see interventions with scientifically proven efficacy, but they have rarely been compared in head-to-head experiments measuring the relative efficacy of different interventions with the same aim. Furthermore, in most cases different preventive interventions are not mutually exclusive, and using interventions together can be a successful strategy [56]. Therefore, the national health system should decide which interventions are better suited to a given context, more acceptable to a given population, sustainable given the available resources, and thus recommended by the National Prevention Plan [32,57]. Having a short list of recommended interventions limits the opportunity to implement a thorough cocreation process. On the other hand, the need to...
apply some criteria to prioritize interventions and adapt the best intervention for a given community among those recommended is a process that can obtain enormous benefits from the application of a cocreation process.

Prevention and health promotion interventions raise another issue: these interventions, particularly those aimed at changing risky behaviors and promoting healthy lifestyles, are not perceived as needs by the target population [58,59]. The target population often considers these interventions unwelcome intrusions in their lives [60]. Furthermore, in prevention, the target population may attribute a value to the possible health benefits of an intervention that is substantially different from that attributed by society (or by public health care professionals). In our specific case (ie, childhood obesity), we observed in our previous studies [5] that many parents were not at all aware that their child was overweight, and that in some cases, parents’ main concern was underweight, even when their child was frankly obese. Again, cocreation tools (in this study, community reporting and focus groups) involving parents who became aware that their children were overweight and health care professionals with experience in counseling these families gave us valuable insight into what is actually perceived as the most important issues in improving the family lifestyles.

When the service provided is a prevention intervention targeting childhood obesity, the definition of users and beneficiaries is also challenging; while the beneficiaries are the children themselves, we are interested in changing the entire family’s behaviors (diet and physical activity are both determined by family habits). Finally, in our pilot project, the users of the app will be only the parents; a positive effect on the beneficiaries (children) can only be achieved by changing the behaviors of the larger target (the family) [61,62].

**Mobile Apps in Health Care Service: Cocreation of a Cocreation Tool**

The next step of the process will be the development of the app prototype during which conversations with the consulting committee subgroups will be still in place. The first app prototype will be tested by users (families and pediatricians). Initially, user feedback will be collected through focus groups. The app itself will collect qualitative and quantitative feedback; the use of different services in the app will be logged and it will be possible to comment on the various app functions by means of a short text or simply clicking on a like or do not like icon.

Cocreation in health promotion and prevention has some unique features that must be taken into consideration. The pilot study we conducted showed that cocreation is feasible. The key co-governance and cocreation tool was the consulting committee. Including stakeholders on this committee made it possible to expand the number of persons and institutions actively contributing to the project.

**Acknowledgments**

We thank Jacqueline M Costa for editing the text.


**Authors’ Contributions**

PGR coordinated the project, in particular the work of the consulting committee. FF and LB were part of the organization of the project, in particular of the consulting committee maintaining contact with all the members. TG was a partner of the project and contributed to the definition of the stakeholder map, and collected the need assessments, trying to transform them for the informatics step. SA, MF, SLS and MES were part of the second and third level of the Bambini Molto in Forma Project. SLS, CDG and MF became community reporters and collect the needs from families. AMF, EF, BI, MT, AMD, CP, AV and SS contributed to the cocreation lab and to the definition of the stakeholder map. CDG and AF contributed to the definition of the stakeholder map and worked on diet and healthy menus. LDA and MF contributed to the information technology development of the app. MP, SS, EF, AMD, CP and AV were part of first level of the Bambini Molto in Forma Project. PGR drafted the article. FF, LB, FV and TG contributed to the methods and results sections of the manuscript and to changes post revision. AB, RP, VM, GG and GM contributed to the draft of all manuscripts. HT developed the part of the article about Community reporting. BI, AMF, SLS, CP and AMD participated in the description of Public Crocation Lab. SA, EF, SLS, BI, CP, SS, MES, AV, FV and AMD performed the literature search and article screening. Members of Working group were part of the consulting committee and worked to find the contents and functions of the app.

**Conflicts of Interest**

None declared.
References


34. Institute of Community Reporters. URL: https://communityreporter.net/ [accessed 2020-04-03]


**BMInforma:** Bambini Molto in Forma

**CoSIE:** Cocreation of Service Innovation in Europe

**IT:** information technology

**LHA:** local health authority

---

Edited by G Eysenbach; submitted 09.09.19; peer-reviewed by S Sankaran, S Smith; comments to author 10.12.19; revised version received 04.02.20; accepted 21.02.20; published 08.06.20.

**Please cite as:**

Describing the Process and Tools Adopted to Cocreate a Smartphone App for Obesity Prevention in Childhood: Mixed Method Study

**URL:** https://mhealth.jmir.org/2020/6/e16165

**doi:** 10.2196/16165

**PMID:** 32357123

© Paolo Giorgi Rossi, Francesca Ferrari, Sergio Amarri, Andrea Bassi, Laura Bonvicini, Luca Dall’Aglio, Claudia Della Giustina, Alessandra Fabbri, Anna Maria Ferrari, Elena Ferrari, Marta Fontana, Marco Foracchia, Teresa Gallelli, Giulia Ganugi, Barbara Ilari, Sara Lo Scocco, Gianluca Maestri, Veronica Moretti, Costantino Panza, Mirco Pinotti, Riccardo Prandini, Simone Storani, Maria Elisabeth Street, Marco Tamelli, Hayley Trowbridge, Francesco Venturelli, Alessandro Volta, Anna Maria Davoli, Childhood Obesity Prevention Working Group. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 08.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Original Paper

User Experiences and Preferences Regarding an App for the Treatment of Urinary Incontinence in Adult Women: Qualitative Study

Nienke J Wessels1, MD; Lisa Hulshof1, MD; Anne M M Loohuis1, MD; Lisette van Gemert-Pijnen2,3, PhD; Petra Jellema1, PhD; Henk van der Worp1, PhD; Marco H Blanker1, MD, PhD

1Department of General Practice and Elderly Care Medicine, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
2Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands
3Faculty of Medical Sciences, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

Corresponding Author:
Nienke J Wessels, MD
Department of General Practice and Elderly Care Medicine
University of Groningen
University Medical Center Groningen
Postbus 196, Huispostcode FA21
Groningen, 9700 AD
Netherlands
Phone: 31 503616746
Email: n.j.wessels@umcg.nl

Abstract

Background: Although several apps are available to support the treatment of urinary incontinence (UI), little has been reported about the experiences and preferences of their users.

Objective: The objective of this study was to explore the experiences and preferences of women using a mobile app for the treatment of UI and to identify potential improvements to the app. We developed this app for three types of UI: stress UI, urgency UI, and mixed UI.

Methods: The participants in this qualitative study were women with self-reported stress UI, urgency UI, or mixed UI who used an app-based treatment to manage their condition for at least six weeks. Following the intervention, semistructured interviews were conducted to explore the participants’ experiences and preferences regarding the app. All interviews were audio-recorded, transcribed verbatim, and analyzed separately by two researchers.

Results: Data saturation was reached after interviewing 9 women (aged 32-68 years) with stress UI (n=1, 11%), urgency UI (n=3, 33%), or mixed UI (n=5, 56%). Accessibility, awareness, usability, and adherence emerged as the main themes. On the one hand, participants appreciated that the app increased their accessibility to care, preserved their privacy, increased their awareness of therapeutic options, was easy to use and useful, and supported treatment adherence. On the other hand, some participants reported that they wanted more contact with a care provider, and others reported that using the app increased their awareness of symptoms.

Conclusions: This qualitative study indicates that women appreciate app-based treatment for UI because it can lower barriers to treatment and increase both awareness and adherence to treatment. However, the app does not offer the ability of face-to-face contact and can lead to a greater focus on symptoms.

JMIR Mhealth Uhealth 2020;8(6):e17114 doi:10.2196/17114

KEYWORDS
ehealth; mobile applications; self-management; qualitative research
Introduction

Approximately one-third of women with urinary incontinence (UI) seek medical attention [1]. In many cases, this low percentage can be explained by feelings of shame and embarrassment, perceptions regarding the normalcy of UI, and beliefs about the treatment options (or lack thereof) for UI. With increasing smartphone ownership, mobile health apps are providing a promising route to improving health care delivery and outcomes [2,3]. It is estimated that half of the approximately 3.4 billion people who use smartphones and tablets worldwide have downloaded a health app [4]. These apps play a particular role in reaching people who suffer from conditions that make them feel embarrassed or stigmatized, and they may help to lower barriers for women with UI [5]. Additionally, it has been shown that mobile app use can increase adherence to treatment advices, thereby improving outcomes and reducing health care costs [6,7]. Although several apps are available to support the treatment of UI, little has been reported about the experiences of their users, which is important for successful implementation of these apps [8]. We developed an app for the treatment of stress UI, urgency UI, and mixed UI. Recently, we have shown that this app is noninferior to care as usual, and usage of the app results in clinically relevant symptom improvement [9]. In the current study, we aimed to explore the experiences and preferences of women regarding the use of this app and to seek their opinions about potential areas for improvement.

Methods

We chose a qualitative design for this study, as this design is especially suitable to explore and inquire into human experiences and preferences. Thematic analyses provide subthemes describing these experiences and preferences, as they are closely interwoven [10]. We conducted semistructured in-person interviews as part of the URinControl study, a mixed-methods study consisting of a randomized controlled trial (RCT) with extensive process evaluation to assess the expectations and experiences of patients and care providers regarding app-based treatment of UI in women [11]. The relevant medical ethics committee approved this study (M17.207954), and the COREQ guideline was followed in this report [12].

Participants

Women were recruited through four primary care practices located in the northern part of the Netherlands. We selected practices that did not participate in the RCT to avoid the possibility of influencing the ongoing trial. We invited women who consulted their general practitioner (GP) for stress, urgency, or mixed UI in the past 10 years; were aged ≥18 years; self-reported UI at least twice a week; wanted treatment; and had access to a smartphone or tablet. The exclusion criteria were indwelling urinary catheter, urogenital malignancy, previous surgery for UI, treatment for UI in the previous year, terminal or serious illness, cognitive impairment or psychiatric illness (reported by their GP), overflow or continuous UI, pregnancy or recent childbirth (<6 months ago), or inability to complete a questionnaire in Dutch. The inclusion and exclusion criteria were identical to the criteria used in the RCT, enabling future triangulation (ie, combination) of the qualitative and quantitative data from the trial, to provide a complete picture of app usage. Purposive sampling was used to achieve diversity in age, UI type, educational level, and geographic location. Interviews were planned approximately 6 weeks after the participants received the login credentials for the app.

The URinControl App

The development of the URinControl app has been described elsewhere [11]. It provides step-by-step advice for treating stress UI, urgency UI, and mixed UI in a patient-friendly format based on the guidelines for treating female UI in primary care [13] (Figure 1).
Figure 1. Representation of the contents of the URinControl app. (A) Information on both types of urinary incontinence, prevention and treatment options, as well as information on the anatomy and function of the pelvic floor. (B) Training programs for both stress and urgency urinary incontinence. (C) Functionalities of the app, including three reminder options, the graph function, and a patient feedback option. UI: urinary incontinence; W.C.: water closet.

Basic Data Collection
After providing informed consent, each participant completed a short questionnaire to record their age, educational status, and duration of complaints. UI type was assessed according to the Three Incontinence Questions (3IQ), a 3-item questionnaire for classifying stress, urgency or mixed UI; mixed UI was further divided into stress-predominant UI, urgency-predominant UI, or mixed UI. To provide a description of the study population, we assessed UI severity with the International Consultation on Incontinence Modular Questionnaire Urinary Incontinence Short Form (ICIQ-UI-SF), a self-completed questionnaire (score 0-21; a higher score indicates greater severity). Furthermore, to measure self-perceived knowledge, comfort, and skill with finding, evaluating, and applying electronic health (eHealth) information to health problems, we used the eHealth Literacy Scale (eHEALS) [14], an 8-item questionnaire (score 8-40; a higher score represents greater literacy).

Semi-Structured Interviews
To develop the semistructured interview guide, we used constructs of the Technology Acceptance Model (TAM) and the Mobile Application Rating Scale (MARS). The TAM is a widely used and validated model in both qualitative and quantitative studies of health apps. It was developed to explain and predict the acceptance and use of technology and is based on two explanatory constructs of users’ adoption: “perceived usefulness” and “perceived ease of use” [15]. The MARS is a validated and reliable scale that was developed to assess the quality of health apps based on four scales: engagement, functionality, aesthetics, and information quality [16].

We also gained insight into the actual app use, progress, and adherence of the participants by analyzing the automatic log data for the user interactions. These data were then used to personalize the interviews, enabling deeper discussion of themes (eg, if a participant had not used certain parts of the app). However, no attempts were made to combine the log data with the qualitative data. The interview guide (Multimedia Appendix 1) was tested in a pilot study before it was implemented.

The first author (NJW), a female physician trained to carry out qualitative interviews, conducted all the interviews. During each interview, the interviewer encouraged participants to talk freely about their experiences and provided opportunities to discuss subjects they felt had not been covered. The interviewer had no prior relationship with the participants. The second author (LH) was also present during the interviews and recorded additional notes (eg, on nonverbal communication).

The interviews took place at each participant’s GP practice, were recorded using a digital voice recorder, and were transcribed verbatim. Respondent validation was used to check the results. At regular points between interviews, we also conducted peer debriefings within the research group to evaluate...
the interviews and interview guide. Interviews were conducted until data saturation was reached, which was defined as the point at which no new codes were identified from 3 consecutive interviews.

**Analysis**

Two researchers (NJW, LH) separately coded the transcripts using NVivo version 11 (QSR International). Thematic analysis was conducted according to the 6 phases of thematic analysis proposed by Braun and Clarke [17]. Data were analyzed by deductive and inductive approaches. The deductive coding framework was assembled using the TAM and the MARS [15] with inductive analysis reserved for responses that were unaligned to this model. After initial coding of each transcript, the researchers convened to compare codes and discuss emerging themes, by which they identified a set of four main themes that captured the essence of the interviews.

**Results**

**Participants**

Data saturation was reached after the ninth interview, after which no new participants were invited, and peer debriefing did not lead to changes in the interview guide. The 9 participants (aged 32 to 68 years) had suffered UI complaints for 18 to 312 months; 3 (33%) had urgency UI, 1 (11%) had stress UI, 2 (22%) had mixed UI, 2 (22%) had urgency-predominant mixed UI, and 1 (11%) had stress-predominant mixed UI. Of the 9 participants, 4 (44%) reported using incontinence pads daily. The median ICIQ-UI-SF score was 8 (range 2-13), indicating low to moderate self-perceived severity of UI. The eHealth literacy scores ranged from 17 to 33, indicating moderate to high self-perceived eHealth literacy (Table 1).

### Table 1. Characteristics of the participants (N=9).

<table>
<thead>
<tr>
<th>No</th>
<th>Age (years)</th>
<th>Educational level</th>
<th>BMI (k/m²)</th>
<th>Type of UI (3IQ)</th>
<th>Duration of UI (months)</th>
<th>Previous treatment for UI</th>
<th>Vaginal deliveries, n</th>
<th>ICIQ-UI-SFd</th>
<th>eHEALSe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>Senior secondary vocational education</td>
<td>28.4</td>
<td>Stress-predominant mixed UI</td>
<td>180</td>
<td>None</td>
<td>1</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>Pre-university education</td>
<td>20.0</td>
<td>Urgency UI</td>
<td>60</td>
<td>Medication</td>
<td>2</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>Bachelor’s degree</td>
<td>19.5</td>
<td>Urgency-predominant mixed UI</td>
<td>36</td>
<td>PPT</td>
<td>0</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>Senior secondary vocational education</td>
<td>36.9</td>
<td>Mixed UI</td>
<td>18</td>
<td>PPT</td>
<td>1</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>Master</td>
<td>32.7</td>
<td>Urgency UI</td>
<td>36</td>
<td>PPT</td>
<td>1</td>
<td>9</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>32</td>
<td>Secondary vocational education</td>
<td>22.2</td>
<td>Urgency-predominant mixed UI</td>
<td>60</td>
<td>PPT</td>
<td>2</td>
<td>7</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>37</td>
<td>Pre-university education</td>
<td>31.0</td>
<td>Urgency UI</td>
<td>72</td>
<td>PPT</td>
<td>0</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td>8</td>
<td>53</td>
<td>Pre-university education</td>
<td>26.0</td>
<td>Stress UI</td>
<td>312</td>
<td>None</td>
<td>2</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>9</td>
<td>42</td>
<td>Bachelor’s degree</td>
<td>25.6</td>
<td>Mixed UI</td>
<td>120</td>
<td>PPT</td>
<td>1</td>
<td>9</td>
<td>30</td>
</tr>
</tbody>
</table>

*Note:*

- BMI (k/m²): kilograms per square meter.
- UI: urinary incontinence.
- 3IQ: 3 Incontinence Questions.
- ICIQ-UI-SF: International Consultation on Incontinence Modular Questionnaire Urinary Incontinence Short Form.
- eHEALS: eHealth Literacy Scale.
- PPT: pelvic physical therapy.

The automatically logged user data indicated that 28-59 days elapsed between app installation and the interview (Table 2), with actual app usage of 2-31 days (ie, the participant opened the app at least once on a day). The highest levels reached among the participants ranged from 0 to 6 out of a maximum of 15 levels.
Table 2. Overview of the log data.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Duration of usage (days)(^a)</th>
<th>Days of usage(^b)</th>
<th>Highest level reached (0-15)(^c)</th>
<th>Use of reminders(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>59</td>
<td>7</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>59</td>
<td>18</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>53</td>
<td>17</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
<td>3</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>49</td>
<td>31</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>95</td>
<td>8</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>74</td>
<td>6</td>
<td>2</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>28</td>
<td>2</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>10</td>
<td>6</td>
<td>Yes</td>
</tr>
</tbody>
</table>

\(^a\)Time in days between app installation and the interview.  
\(^b\)Days in which the app was opened at least one time.  
\(^c\)The highest level opened at least one time.  
\(^d\)Yes: turned on reminders at least once.

**Interviews**

The interviews lasted 15-37 minutes, and thematic analysis of the transcripts resulted in four main themes concerning the experiences and preferences of women with the URinControl app (Table 3). These themes were accessibility, awareness, usability, and adherence.
Table 3. Participants’ themes presented as main themes, subthemes, and meaningful codes.

<table>
<thead>
<tr>
<th>Main themes and subthemes</th>
<th>Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>Available 24 hours, 7 days a week</td>
</tr>
<tr>
<td></td>
<td>Convenient</td>
</tr>
<tr>
<td></td>
<td>Easily accessible</td>
</tr>
<tr>
<td>Privacy</td>
<td>Keeps symptoms private</td>
</tr>
<tr>
<td></td>
<td>Lowers the barrier to talk about UI</td>
</tr>
<tr>
<td></td>
<td>Hinders privacy</td>
</tr>
<tr>
<td>Awareness</td>
<td>Confrontational</td>
</tr>
<tr>
<td></td>
<td>Worsening of symptoms</td>
</tr>
<tr>
<td></td>
<td>Coping strategies</td>
</tr>
<tr>
<td></td>
<td>Treatment options</td>
</tr>
<tr>
<td></td>
<td>Effect of therapy</td>
</tr>
<tr>
<td></td>
<td>Nothing new</td>
</tr>
<tr>
<td>Usability</td>
<td>Feeling familiar with app</td>
</tr>
<tr>
<td></td>
<td>Clear</td>
</tr>
<tr>
<td>Usefulness</td>
<td>Credible source</td>
</tr>
<tr>
<td></td>
<td>Functionalities</td>
</tr>
<tr>
<td></td>
<td>Graphs unclear</td>
</tr>
<tr>
<td></td>
<td>App with therapist</td>
</tr>
<tr>
<td></td>
<td>Information for health care provider</td>
</tr>
<tr>
<td></td>
<td>App without therapist</td>
</tr>
<tr>
<td>Adherence</td>
<td>Too busy</td>
</tr>
<tr>
<td></td>
<td>Interference of daily activities</td>
</tr>
<tr>
<td></td>
<td>Fit into daily routine</td>
</tr>
<tr>
<td></td>
<td>Privacy</td>
</tr>
<tr>
<td></td>
<td>Improvement</td>
</tr>
<tr>
<td></td>
<td>Progress in levels</td>
</tr>
<tr>
<td></td>
<td>Reminders</td>
</tr>
<tr>
<td></td>
<td>Motivational feedback</td>
</tr>
</tbody>
</table>

**Theme 1: Accessibility**

Accessibility concerns the women’s experience with access to UI health care with the app and the associated privacy. Most participants considered that the app was convenient and improved access to UI care and treatment. Information about the exercises was always available; therefore, the women could perform them when they wanted. Using the app was also considered less burdensome than receiving treatment from a health care provider:

*You always have to make an appointment to see your physiotherapist. And everyone’s so busy these days that you think: why bother? And this gives you the support you need, because in theory it’s available 24 hours a day. And I really liked that.* [Participant 7]

Keeping UI symptoms private was important to some participants. They expressed that they found it difficult to discuss their symptoms with others and that they dreaded the internal examinations often performed by care providers. The accessibility of the app made it easier to receive treatment while keeping their complaints private:

*I have control over the app, and I can keep it private. I find it difficult to talk about this with someone (and it definitely has to be a woman, not a man) and make...*
Theme 2: Awareness

Participants reported gaining greater awareness of their symptoms and of potential coping and therapeutic strategies. Most participants had become used to their UI symptoms and had learned to live with them, and using the app increased their awareness. For some, this led to a realization that their symptoms were worse than they had previously thought:

"... but I was also triggered by it. I thought: wow, this is really much worse than I thought, because I always adapt to this incontinence [...]" [Participant 2]

Two participants with urgency UI experienced a worsening of symptoms after shifting their focus; however, others felt they became more aware of their dysfunctional coping strategies and how they could be addressed:

"I just got even more incontinent; it became even more unpleasant." [Participant 5]

"I spoilt myself really by going to the toilet when I had to, and I had to learn not to." [Participant 2]

Women became more aware of the conservative treatment options for UI, and in some instances, they changed their beliefs about the possible effects of these treatments. One woman with urgency UI who had received treatment for UI in the past was surprised by the impact of conservative treatment on her complaints:

"If you persist and are consistent, there’s more you can do about it than I initially thought was possible. I was convinced that I would have to be operated on, but I find I’ve made a lot of progress, so I think if I carry on for a little longer, it can even get much better." [Participant 7]

Several participants realized that to relieve their symptoms, they had to do the exercises for longer than they had initially hoped:

"I have less difficulty now in making it to the bathroom, but I noticed that you have to train longer than you actually would want." [Participant 7]

Unfortunately, the exercises only repeated previous training experiences for women who had already undergone pelvic floor muscle training; some of these participants had hoped that the app would offer new options.

Theme 3: Usability

Participants from various ages and educational levels stated that the app was easy to use, noting that it was self-explanatory and that they became familiar with it quickly. They appreciated its simplicity, clarity, and visual appeal:

"The app was very clear, really easy to understand." [Participant 9]

The women were positive overall about the usefulness of the app and stated they would recommend it to others. Concerning the different app functions, the information provided in the app was noted to be clear and useful. This was especially so for women who had not been treated for UI previously; in contrast, those who had already undergone pelvic physical therapy reported that it was repetitive. Some participants would have preferred less text with fewer difficult words. One participant appreciated that the app provided a credible source of information:

"Of course, you can find films on YouTube, but you can never be sure whether they’re the right ones, because there’s a lot of rubbish on the internet. So that’s what I really like about it." [Participant 7]

Participants appreciated the instructional videos in the app and found them more appealing and motivational than pictures. One participant reported that videos made it feel as if she was doing the exercises with someone else:

"The videos are definitely useful. They show you how to exercise more efficiently. That makes it less a case of: [makes a sound expressing disgust] I have to exercise again. You just sit down and get on with it." [Participant 9]

The distraction games included in the treatment program for women with urgency UI were used infrequently; most participants wanted greater variety or challenge to keep them engaged.

Concerning data presentation, some women found the included graph function to be unclear and stated that they did not understand how to interpret the graphs. Some merely glanced at the graphs, but others appreciated the statistics:

"Yes, I think it’s interesting to look at it that way [the graphs]. [...] You just can’t cheat." [Participant 3]

There were also comments regarding input from health care providers. Several mentioned that the support and guidance of a pelvic physical therapist could add value, especially for older women who might have trouble navigating the app. One woman stated that face-to-face contact for ongoing training and support could improve her adherence. In other cases, participants suggested that information from the app could be used to inform their health care providers. A woman who had previously attended pelvic physical therapy reported that she missed the direct feedback, stating that the app would have been a valuable addition to that therapy:

"I think it would be a good addition to a pelvic floor specialist, who does the exercises with you and lets you feel which muscles you have to tense. Because I really missed that." [Participant 5]
Other women preferred to use the app alone because of poorly defined barriers to visiting pelvic physical therapists:

I feel there’s something holding me back from going to the physiotherapist specifically for that [the incontinence]. [Participant 6]

Theme 4: Adherence

Most participants experienced adherence and motivation issues; they found it difficult to perform exercises three times a day because they often felt they were too busy with daily activities. These women typically had trouble prioritizing treatment:

I’m always busy with the kids. [Participant 6]

[About the app’s strengths] ‘Well, that it’s quick and easy to do. But not if you’ve got a busy job, or a very busy private life.” [Participant 8]

The need for privacy also played a role when trying to fit the exercises into daily routines:

I actually wanted to do them [the exercises] quite frequently when I was on my own. So, when my children were in bed or after I’d taken them to school I thought: now, I’ve got all the time for it. So, I made it part of my own personal ritual. [Participant 7]

Participants mentioned that noticing symptomatic improvement motivated them to persevere. One stated that being able to see progress in the exercise program motivated her, and another was motivated by the challenge to reach the highest level:

I still use it [the app] and I’m going to carry on with it. I’m a bit of an overachiever, so I want to get to the highest level. [Participant 9]

Most of the participants stated that the inclusion of automated reminders was an important feature. However, although reminders were set by the participants, one woman commented that the timing was not always convenient:

At least for me, it [the reminder] acts as a trigger, like, oh yes, it’s that time again. [Participant 3]

You get reminders to do these exercises, I liked that. And I made use of that option. But if you’re doing the washing up at that moment, or walking the dog, then you just click it away and an hour later you’ve forgotten about it. [Participant 4]

Finally, one participant mentioned that she wanted feedback that was more stimulating and provided more praise to keep her motivated and engaged.

Discussion

eHealth is an emerging area of health care technology with an ever-growing number of apps to assist and monitor patients with various health complaints. Research into the effectiveness and user experiences of these health apps is increasing. However, studies investigating apps for the treatment of UI are limited. We therefore aimed to explore the experiences and preferences of women regarding the use of an evidence-based mobile app that we had recently developed for the treatment of UI.

Literature on this topic is scarce. A review of the expectations and experiences of women concerning eHealth applications for UI [18] identified only one study of an online self-help program with email support of stress UI [19]. Recently, studies have been published on eHealth for women with stress UI, including their expectations of eHealth [20] and their experiences of an app focused on pelvic muscle training [21]. Notably, the two studies on experience with eHealth identified themes that overlapped considerably with those in our study [19,21].

Principal Results and Comparison With Prior Work

Asklund et al [21] identified enabling my independence as a core theme reflecting desire to manage incontinence independently, with three subthemes of something new!, keeping motivation up!, and good enough? The accessibility and something new! themes are both related to accessibility of care, and the themes adherence and keep motivation up! both relate to treatment adherence. There is also overlap between the usability and good enough? themes; both involve uncertainty about performing exercises correctly without supervision. Björk et al [19] also identified a theme of hidden but present, which is closely related to our accessibility theme. However, neither of the earlier studies specifically identified an awareness theme, though Björk et al mentioned that participants had increased awareness of how to handle their incontinence. Given that the theme of increased symptom awareness was absent from earlier research and mainly arose from interviews of women with urgency UI, it may be a specific concern in this cohort.

The participants valued the easy access to the health care app throughout the day. Issues with privacy and talking about UI problems were raised by the participants, who indicated that they were attracted to the idea of receiving health care without consulting a health care provider. The app also made it easier for some women to talk about their complaints with friends or relatives. This is similar to the finding of Björk et al [19], who reported that treatment for stress UI without face-to-face contact could break down barriers such as shame. However, some women still wanted to keep their symptoms private and preferred to use the app without consultation or care provider support. In contrast, other participants wanted greater care provider input from the start, particularly to confirm correct exercise performance and to promote adherence. These latter findings are consistent with studies in other fields indicating that participants prefer face-to-face contact in eHealth treatments [18,22].

Although other studies of UI apps have not reported greater self-awareness following app use, this outcome has been reported in studies of depression and chronic obstructive pulmonary disease [22,23]. Two of our participants expressed that greater self-awareness could be negative because they assessed the increased awareness as unpleasant and confrontational. The app treatment for urgency UI starts with monitoring toilet visits by pushing the “pee-button” and by advocating distraction techniques to extend the intervals between toilet visits. This treatment required these women had to change advocating distraction techniques to extend the intervals between toilet visits. This treatment required these women had to change their behaviors and could induce feelings of insecurity and self-blame, especially if they were unsuccessful. Increased symptom awareness could therefore explain the worsening of
symptoms. However, for other women, the app increased their awareness of the available treatment options and even changed some beliefs about the efficacy of conservative treatment. Several studies have shown that women are not always aware of treatment options or their effectiveness, which can adversely influence health-seeking behavior [1,24].

Perceived ease of use and perceived usefulness are the most important predictors of technology acceptance according to the TAM [15]. Women perceived our app to be both easy to use and clearly structured, and they appreciated that it was a credible information source, as shown previously [25]. However, while some women found the included information to be useful and easy to read, others wanted simpler text. This finding is consistent with research by Peng et al [26] and suggests the need for tailored information based on user preference.

Ensuring and monitoring treatment adherence is a major challenge when managing UI [27], and research has shown that adherence to eHealth apps is low in the absence of face-to-face contact [18]. Although the automatically logged user data showed low app use, this does not necessarily reflect true treatment adherence. Indeed, many women will undoubtedly perform the exercises without opening the app each time, especially as they become more practiced. Some participants would have liked more features to keep them engaged with the app, such as better feedback and a greater variety of distraction games. The importance of features supporting app engagement is stressed by studies showing that repeated use of health apps over time is low despite their wide availability [28,29]. The greatest barrier to adherence in this study was that the participants were too busy. Participants stated that they regularly forgot to perform the exercises and appreciated the reminder function, which was supported by the usage levels in the log data. Consistent with this finding, the importance of reminder functions in eHealth tools has been stressed in previous reports [23,25,26].

Strengths and Limitations
We are not aware of any prior study of patient experience with mobile treatment for all three types of UI. To improve the robustness of our research, we integrated constructs of existing theoretical frameworks (TAM and MARS) in the interview guide, used log data to personalize the interviews, and confirmed the results by respondent validation. Moreover, we deliberately included a heterogeneous cohort based on age, level of education, and type of UI. However, women were recruited through only four practices in the northern part of the Netherlands. Therefore, the results may not be transferable to other GP populations. Moreover, although we analyzed the experiences of women with our app-based treatment over 6 weeks, we did not explore their experiences in the long term. Furthermore, although there were some indicators of differences in experiences between the UI subgroups (eg, increased symptom awareness led to worsening of symptoms for some women with urgency UI), the design of this study did not allow further analyses of the differences between these subgroups.

Conclusions
In this qualitative study that incorporated log data, we showed that women tended to appreciate the URinControl app for treating UI. Use of this app may lower barriers to seeking treatment, increase self-awareness, and support treatment adherence. However, some women wanted more information about new therapies, more variety in the distraction games, contact with care providers, and improved feedback. Others wanted simplification of some areas, such as text with less detail and complexity and more understandable graphs. Notably, several women experienced a negative impact as awareness of their symptoms increased. These points of improvement will be taken into account with further development of the URinControl app. Furthermore, the experiences and recommendations outlined in this study can be used to optimize the implementation of this app in the future.

Acknowledgments
We thank the participants for their time and effort, Dr Robert Sykes (www.doctored.org.uk) for providing editorial services, and Kumar Jamdagni for translating the citations.

This work was supported by a grant from ZonMw, The Dutch Organisation for Health Research and Development (project number 837001508), and subfunded by a grant from the P.W. Boer Foundation. This study won the 2016 Professor Huygen Award for best study proposal in general practice, which resulted in additional funding. The funders had no role in data collection, data analysis, the decision to publish, or the manuscript preparation.

Authors’ Contributions
All authors had full access to all the study data and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest
The authors were involved in the development of the URinControl app.

Multimedia Appendix 1
Interview guide.
[DOCX File, 20 KB - mhealth_v8i6e17114_app1.docx]
References


Abbreviations

3IQ: 3 Incontinence Questions  
eHEALS: eHealth literacy scale  
GP: general practitioner  
ICIQ-UI SF: International Consultation on Incontinence Modular Questionnaire Urinary Incontinence Short Form  
MARS: Mobile Application Rating Scale  
PPT: pelvic physical therapy  
RCT: randomized controlled trial  
TAM: Technology Acceptance Model  
UI: urinary incontinence

©Nienke J Wessels, Lisa Hulshof, Anne M M Loohuis, Lisette van Gemert-Pijnen, Petra Jellema, Henk van der Worp, Marco H Blanker. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 12.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Mobile App Adopting an Identity Focus to Promote Physical Activity (MoveDaily): Iterative Design Study

Floris Hooglugt¹, MSc; Geke D S Ludden¹, MSc, PhD
Department of Design, Production and Management, University of Twente, Enschede, Netherlands

Corresponding Author:
Geke D S Ludden, MSc, PhD
Department of Design, Production and Management
University of Twente
Drienerlolaan 5
Enschede, 7522NB
Netherlands
Phone: 31 534892920
Email: g.d.s.ludden@utwente.nl

Abstract

Background: Web-based and mobile interventions to influence physical activity behavior have had limited effects on sustained behavior change. One reason may be that the interventions aim to change largely habitual behavior. Following an identity-oriented approach could be a successful strategy to behavior change because people are committed to behave in line with their self-perception of identity.

Objective: In this paper, we take a closer look at the role of motivation in long-term adherence to lifestyle interventions. The paper outlines a method for web-based or mobile intervention development that allows exploration of integrating behavior change theory into the design process. We will describe the development of a mobile app that allows people to be self-determined and to value and self-regulate physical activity by adopting an identity-oriented approach.

Methods: This paper describes a Research through Design (RtD) process in which design activities are carried out as part of the knowledge-generating process. Two RtD phases were completed, followed by a conceptual design phase. In the first RtD phase, 8 participants used diary cards to study initial attitudes toward starting with small changes in physical activity. In the second RtD phase, 26 participants used a web-based app to study changes in physical activity. We used an adapted version of the Self-Report Habit Index (SRHI) to evaluate individuals’ perceptions of a particular behavior with respect to the three facets of a habit. The conceptual design phase consolidated the results from first two RtD phases into a design of a mobile app that combines an identity approach with gamification principles. The conceptual design was evaluated in a user-experience study with 4 participants.

Results: In the first RtD phase, we found that interacting daily with diary cards and reflecting on physical activity patterns is a promising strategy but works better through a digital medium. In the second RtD phase, SHRI ratings from all participants generally increased each week. In the conceptual design phase, we found that the concept of the mobile app was positively evaluated by participants. However, participants mentioned that terms such as “identity” do not resonate with them and that scenarios could be simpler.

Conclusions: This paper provides deeper insights into designing for electronic health (eHealth) interventions and services and suggests a new way that motivation can be shaped by the design of an intervention and adherence to physical activity. To the best of our knowledge, this was the first iterative design study in which the effects of adopting an identity approach to both motivation and physical activity were included and observed. Initial promising results were found for using a web-based intervention where habits and identification with the personal importance of a behavior were repetitively triggered.

(JMIR Mhealth Uhealth 2020;8(6):e16720) doi:10.2196/16720

KEYWORDS

research through design; physical activity; habits; identity; behavior change; mHealth; design
In this paper, we take a closer look at the role of habits and motivation in long-term adherence to lifestyle interventions. We will explore how the design of a mobile app can incorporate a focus on self-directed motivation. A self-determined form of motivation is regulation through identification. Over time, identified regulations can be fully adopted by a person as belonging to him or her [29]. Identification occurs when the person has identified with the personal importance of a behavior and has, thus, accepted its regulation as his or her own. An individual who always takes the stairs instead of the elevator because she or he sees it as relevant to her or his health, which she or he values as a life goal, has identified with the value of this physical activity [30]. Studies in relation to smoking behavior have shown that stronger smoker self-identity (ie, thinking of the self as a person who smokes) predicts fewer quitting attempts [31,32]. People are committed to behave in line with their self-perception of identity and, therefore, behavior change and identity change depend upon each other [33].

**Background**

As a response to the worldwide increase in so-called lifestyle diseases such as obesity and diabetes [1], efforts to raise people’s awareness of the importance of living a healthier life, as well as to motivate and support people to make lifestyle changes, have widened from traditional information campaigns to monitoring and coaching systems [2,3]. As a result of this, we have seen a multitude of web-based and mobile interventions as well as dedicated trackers, designed to support people in adopting a healthy lifestyle, reach the market [4,5]. Interventions are typically aimed at adopting a healthier diet or becoming more physically active and, in some cases, a combination of these two.

Although multiple studies [6,7] describe the effectiveness of web-based and mobile interventions on influencing physical activity behavior, others describe the uncertainty of their efficacy [8]. To date, it remains unclear which behavior change techniques lead to greater intervention success in increasing physical activity [7,9-14]. In addition, while a wide range of behavior change interventions exist that use similar techniques, studies have concluded that these generally lack the use of theoretical constructs [15-17]. It is worth noting that at the same time, we know that successful interventions are informed by grounded behavior change theory [18,19]. Furthermore, adherence to an intervention is essential for a positive effect on health to transpire [20], yet web-based and mobile interventions have not been as successful as typically traditional individual and or group treatment approaches for long-term adherence [21-23].

It seems difficult to single out one specific reason for the lack of success of interventions in achieving long-term behavior change. Several researchers have proposed that the affective experience of persuasive technologies is the key to their effectiveness [24,25]. In other words, a better design that is not only functionally effective but also desirable and engaging could improve acceptance of, and thereby adherence to, interventions [20]. Another reason for the lack of success may be that interventions try to target behaviors that are largely habitual in nature [23], often occurring outside the conscious awareness of people [26]. Many energy balance-related behaviors (eg, unhealthy eating) are performed habitually, with little forethought [27]. Interventions that focus on breaking these established habits will, therefore, face difficulties in the long term because people behave according to their habits even when motivated not to do so [28].

In this paper, we take a closer look at the role of habits and motivation in long-term adherence to lifestyle interventions. We will explore how the design of a mobile app can incorporate a focus on self-directed motivation. A self-determined form of motivation is regulation through identification. Over time, identified regulations can be fully adopted by a person as belonging to him or her [29]. Identification occurs when the person has identified with the personal importance of a behavior and has, thus, accepted its regulation as his or her own. An individual who always takes the stairs instead of the elevator because she or he sees it as relevant to her or his health, which she or he values as a life goal, has identified with the value of this physical activity [30]. Studies in relation to smoking behavior have shown that stronger smoker self-identity (ie, thinking of the self as a person who smokes) predicts fewer quitting attempts [31,32]. People are committed to behave in line with their self-perception of identity and, therefore, behavior change and identity change depend upon each other [33].

**Habits and Identity**

By experiencing positive feedback on daily habits, new beliefs about one’s identity can be formed. We visualized the relationship between habit forming and identity as two overlapping loops: the habit loop and the identity loop (see Figure 1). Any habit follows a closed loop of trigger, action, and reward [34,35]. The habit loop is part of the identity loop, showing how behavior that is performed consistently forms the starting point of changing one’s beliefs. Concurrently, the identity loop visualizes how beliefs can, in turn, influence behavior [33]. The identity loop has four phases (see Figure 1): behavior, experience, identity, and expectation. Following a performed activity (ie, the habit), an experience will cause a belief about one’s identity to be formed or reaffirmed. Out of that belief, an expectation about future behavior takes shape. A positive experience will allow for individuals to become more self-determined about the activity and for internalization to occur. Subsequently, the expectation of successfully performing the habit the next time becomes more realistic. If the behavior is repeated, it will lead back to a similar experience or result, which will reinforce one’s beliefs again. It is conceivable that, initially, these identity-focused habits will leverage extrinsic motivation via identification and progress over time to self-regulated integration.

By exploring the development of habits and a change in thinking about identity, we intend to find out how the design of a mobile app might support a sustainable increase in physical activity. The intervention focuses on creating habits through starting small, following the idea that if one’s ability is high (ie, by starting with a simple activity), motivation can be low, yet an individual will still be able to perform the behavior [36].
Objectives
The primary aim of this paper is to outline a method for web-based or mobile intervention design and development that allows for integration of behavior change theory in the design process. We will describe the development of a mobile app that allows people to be self-determined and to value and self-regulate physical activity by adopting an identity-oriented approach. The design process and the iterative evaluation of this app called MoveDaily can serve as an example of a holistic approach to design for behavior change. In this way, this paper contributes to knowledge on the value of adopting an identity approach to behavior change interventions aimed at increasing physical activity. Furthermore, it describes an iterative and user-centered Research through Design (RtD) process [37-39], thereby adding to the body of knowledge on how designers can work with behavior change theory and strategies while designing for behavior change. The iterative evaluation and experimentation with the prototypes will provide knowledge about how features of the design influence people’s motivation and behavior. We will discuss both insights gained through the design activities and those gained through evaluation of the effects of the early prototypes. Together, they will contribute to insight on how people respond to an identity approach to behavior change. Our RtD process is comprised of three phases; we will discuss methods and results in the corresponding sections for each of the three phases and end with a general discussion of the insights and results obtained throughout the three phases.

Methods
Research Through Design
In an RtD process, design activities are carried out as part of the knowledge-generating process. They often include the development of early prototypes that could be mistaken for a product but that are specifically designed to generate knowledge. The process of translating the abstract concepts of identification into designed elements or features of an app is considered an important part of the work [39]. Our RtD process consisted of two main phases that were followed by a conceptual design phase (see Figure 2). The fidelity of the created prototypes increased with each sequential phase. During the first phase, participants’ initial attitudes toward a paper intervention (ie, diary cards paper prototypes), in which they started with small physical activities, were observed. During the second phase, participants’ attitudes toward a web-based intervention (ie, web-based prototype) were studied, as well as participants’ needs and desires for longer-term use. The following design phase described the conceptual design of the mobile app MoveDaily.
Phase 1: Diary Cards Prototypes

Overview
The aim of this phase was to explore how to perform a first translation from the framework in Figure 1 to a design and to obtain insight into initial attitudes of participants about increasing their physical activity.

In this first RtD phase, we studied initial attitudes toward starting small and whether this attitude could change over time. During this phase, we evaluated four intervention types over a 4-week period. The interventions’ features changed every week and were designed with the aim to learn about participants’ attitudes and opinions. The first three interventions were designed as diary cards that changed every week and that were given to the participants daily as reminders of being physically active. Every week the intervention was adjusted to better fit the needs of the participants. The diaries were designed in a way that allowed for a simple and pleasurable interaction, while also being easily adjustable for each week’s variation; see Figure 3 for an example of the diary cards. In the first week, diary cards were aimed at achieving a mindset change in regard to starting small. In the second week, the cards were aimed at supporting participants to determine for themselves at what rate to increase the amount of activity of their routine. In the third week, participants were expected to start enjoying their small activities. In the fourth week, a digital form of the intervention was introduced to evaluate whether using the intervention via a digital medium was preferred over paper.
Participants

A small number of participants was involved because the aim of this phase was to explore what would be the best format of the intervention to integrate the identity-habit loops. Recruiting participants was done internally within a company through announcements via the digital medium Slack. Participants included 8 native Dutch-speaking employees: 5 women (63%) and 3 men (38%) in their 30s and 40s.

Protocol

Before starting the trial, participants chose between three different identities and were asked to select the one that they would like to adopt. The options were as follows: “I am a person who walks” (walking), “I am a strong person” (strength), and “I am a stairs person” (taking the stairs).

Participants reported to the researcher about their physical activity daily. Additionally, short interviews were held weekly to understand each participant’s attitudes toward that week’s diary cards. The discussion topics were partly based on what participants reported on their diary cards, but questions were mostly asked about how easy it was for each participant to perform the physical activity (i.e., automaticity for an indication of habit strength) and if they wanted to increase the difficulty of the physical activity. When a participant decided to discontinue the trial, an exit interview would take place. This interview was mainly aimed at gathering feedback from the participants on how to improve the intervention.

Phase 2: Web-Based Prototype

Overview

In this second RtD phase, we studied participants’ reactions and behavior using a simple web-based intervention. Lessons learned in the first RtD phase were incorporated into a digital prototype in the online survey software Typeform. The main feature of the web-based prototype was to present users with a daily reminder of the identity they selected (see Multimedia Appendix 1). The intervention was presented to participants as an actual service named MoveDaily. It has a bold and playful identity, with fitting color schemes (see Figure 4). Both qualitative and quantitative data were collected to study the participants’ views and attitudes as well as their self-reported behavior. Specifically, we used both a questionnaire and follow-up interviews. Following Schmidt and Retelsdorf [40] and Verplanken et al [41], an adapted Self-Report Habit Index (SRHI) model [42] was used in the questionnaire to report on habits regarding physical activity (SRHI-P) (see Multimedia Appendix 2). Responses were made on 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).
Participants who continued to use the intervention were expected to strengthen their habits over time, meaning the longer the participants adhered, the more consistent their physical activities would be. First, a high SRHI-P score was expected to correlate with a high score on physical activity frequency. Second, a high SRHI-P score was also expected to be associated with more positive remarks by the participants in the interviews and questionnaires regarding the web-based intervention. Third, positive correlations between whether participants recognized feeling free of any pressure or performance and the amount of physical activity performed were expected. Fourth, positive correlations between early dropouts and negative remarks about the intervention (e.g., not fun) were expected. Fifth, a stronger physical activity habit should not only lead to more physical activity but also to a stronger personal image as someone who is comfortable with physical activity. Therefore, positive correlations with physical activity achievement and taking up extra movement were expected.

**Participants**

There was a total of 26 participants, including 22 native Dutch-speaking university students (85%), 1 Portuguese student (4%), 1 French student (4%), and 2 Dutch working people (8%); the sample was made up of 15 women (58%) and 11 men (42%), aged 17-63 years (mean 25 years, SD 4). Participants were recruited by making use of a convenience sample, in which the subjects were selected because of their convenient accessibility and proximity to the researcher, and snowball sampling, which made use of recruitment of participants by other participants already in the study (i.e., the previous RtD phase). Psychology students were ineligible for the study to ensure questionnaire interpretation could not be biased by psychology training. This study methodology does not impose direct sample size constraints; therefore, no cap was placed on the number of participants. A sample of 20 participants was deemed likely to capture a broad-enough range of viewpoints and problems and was, therefore, the goal. We were able to recruit 26 participants.

The testing period was 10 weeks in length: 50 habit-forming days and 20 rest days.

**Protocol**

The online platform Typeform was used for the intervention and as the main data-gathering tool. Typeform is a web-based platform for collecting and sharing information, in a “conversational, human way.” It allows for the fast adjustability that was necessary, as the MoveDaily prototype needed to be adjusted daily in order to appear flexible and smart. Typeform was also capable of saving data input from participants. Email and WhatsApp were adopted as the main tools used for daily interaction with the users, while interviews were done through phone calls.

To start using the intervention, participants signed up by completing an 18-item questionnaire to gather demographics and to allow them to choose between three behaviors: taking a walk, climbing the stairs, and doing exercises. Each participant selected one behavior as the focus for the intervention. A short tutorial of the intervention followed, and a video explained how starting small can be promising.

Participants received daily messages via email at the time specific to their exercise time. Furthermore, participants reported about their activities daily through Typeform. Every week a random selection of 2-3 participants was interviewed about their experiences and the interview data were manually noted down. If participants decided to discontinue using the intervention, they were asked to fill out a dropout form. After participants filled out this form, a follow-up interview was held to confirm understanding of the reason for dropout.

**Phase 3: Conceptual Design**

**Overview**

By consolidating the results from RtD phases 1 and 2, a conceptual design for the MoveDaily app was made. MoveDaily is an identity-oriented app supporting sustainable habits around
physical activity. The app is shaped like a simple explorer game, leveraging gamification as a means to achieve initiation and retention of desired behaviors [43-45]. The conceptual design of MoveDaily was used in a usability and user-experience evaluation; the interactive prototype can be accessed online [46].

Before designing MoveDaily, we specified what the conclusions of the two RtD phases meant for the design of a mobile app. This led to the following five design guidelines:

1. Activate the user from the start by designing the habit during the first time of use.
2. Create a performance-free product, stimulate physical activity without introducing the notion of sport or exercise, and emphasize the benefits of starting small.
3. Increase understanding about habit forming. By doing so in a playful manner, people can start to understand how their healthy behavior can be strengthened and why we start with small activities.
4. Change users’ mindsets from goal oriented to identity oriented.
5. Design for daily interaction.

In **Figure 5**, an example of a feature derived from the design guidelines is shown, where the user is prompted to reaffirm to themselves that they can be a certain kind of person. Due to this self-reevaluation after the activity, people assess how they think and feel about themselves with regard to the performed behavior and create a new self-image (eg, Prochaska et al [47] and Velicer et al [48]). In addition to the five design guidelines, knowledge on successful persuasive features informed the design [49]. An example of such a feature is a visual representation of a user’s progress for positive reinforcement (see **Figure 6**).

**Figure 5.** A reaffirmation example of the app MoveDaily. If an activity was performed successfully, the user is complimented and reminded of his or her daily steps forward.
To come to a high-fidelity conceptual design, we started by sketching variations of app structure, interactions, and layout on paper. Following this phase, wireframes were created in Sketch, a digital design toolkit (see Figure 7). Wireframes allow the designer to further experiment with ideas and concepts on a higher-fidelity level and allow for checking the interaction flow. Next, the wireframes were visually designed and an interactive prototype was created in the workflow tool InVision (InVisionApp Inc).

Figure 6. Two examples of the app MoveDaily as to how progress can be visualized: through a level indicator of the avatar (image on the left) and through a timeline showing all performed activities (image on the right).

Figure 7. Selection of wireframes informing the final design of the mobile app MoveDaily.

Participants
Participants were recruited by making use of convenience sampling. We recruited 4 participants, including 2 native Dutch-speaking university students (50%) and 2 Dutch working people; the sample was made up of 2 women (50%) and 2 men (50%), aged 22-31 years. Besides no prior experience with the project, no exclusion criteria were used for participant selection.
Protocol
The interactive prototype of MoveDaily was installed on an iPhone 6 device and the prototype was presented to participants during a single 45-minute session. After presenting a use scenario, participants were asked to interact with the prototype. Four scenarios were provided sequentially: (1) first-time use, (2) being successful in performing the daily habit, (3) being unsuccessful in performing the daily habit, and (4) personalizing the experience. After each scenario, questions were asked regarding the experience with the app. Following usability testing guidelines [50], the interviewer was reminded to keep neutrality in response to user comments and behavior and read the script to each user in the same way. A notetaker observed and kept track of user comments and behavior. We relied on the qualitative analysis of user feedback for this study.

Results
Overview
To explore how people respond to an identity approach to behavior change, we adopted an iterative design approach of three phases, increasing the fidelity in each sequential phase. In this section, we summarize the results of these three phases, specifically the self-reported data of the diary cards prototypes, self-reported feedback through the web-based prototype, the follow-up interviews, the conceptual design of the MoveDaily app, and the result of a usability test.

Phase 1: Diary Cards Prototypes
A total of 8 participants took part in the diary cards study. Out of 8 participants, 2 (25%) dropped out within the first week. Out of the 6 remaining participants, 1 (17%) chose to leave the study at one point during the 10 weeks. A total of 4 participants from the original 8 (50%) opted for the walking program, while the other 4 (50%) opted for the exercise (ie, strength) program. No participant opted for the stairs program; most participants found this exercise to be too tiring. Transcripts of the interviews were made and these were analyzed by tallying the most common topics and themes.

From the self-reported data, we can observe that participants generally liked how they were conscious of their daily activity through the diary cards (see Table 1). While some described them as practical and simple, one participant also mentioned in one of the early interviews, “It feels like the cards attempt to put an application’s functionalities within paper. I would prefer digital reminders, now I have to be aware of the piece of paper.” The first prototype was not favored as it took too long to complete, nor was it experienced as fun. A total of 5 participants out of 8 (63%) struggled remarkably more than the others with remembering to fill in their diary cards. Cards that had a shorter completion time and features enhancing pleasure, such as rewarding yourself with a sticker, were evaluated more positively. While the digital medium used in the fourth week was limited in its possibilities, it did confirm the benefits of a digital medium versus a paper-based intervention. As expected, initial reactions of participants toward starting small were negative. Participants initially wanted to start with, for example, 40 squats. Surprisingly, all participants increased the difficulty of their exercise only by small margins weekly.

Phase 2: Web-Based Prototype
In phase 2, 26 participants were included. We observed diverse reactions to the web-based intervention. Out of 26 participants, 5 (19%) dropped out within the first week. A total of 10 of the 21 remaining participants (48%) decided to leave at a later time during the study, citing various reasons. Furthermore, while some participants successfully performed and maintained their activity from the beginning, others were unable to consistently commit to the intervention. We included data received from participants who dropped out before the end of the study in our analyses.

To study the participants’ responses from the self-report and interview data, the most common topics and themes were tallied (see Table 1). Results from analysis of this data indicated that the MoveDaily concept was received positively by the participants. While not all participants seemed to understand the concept fully, almost all would mention some key aspects of MoveDaily, such as “habit forming,” “learning a new behavior,” “free of pressure,” “starting small,” or “personal.” It was difficult for the participants to specifically indicate how these concepts became apparent to them, but answers such as “the message on the bottom of the mail,” “the introduction movie,” or “the whole thing” were frequently given.

To evaluate whether people rated the strength of their habits higher over time, a linear regression analysis was calculated to predict SHRI-P based on time (ie, week). A significant regression equation was found ($F_{1,7}=9.644, P=.02$) with an $R^2$ of 0.58. A more detailed inspection of the data did reveal that there were large differences in the way participants assigned scores to rate the strength of their habits. The SRHI-P scores from the first week, for example, ranged from 1.5 to 5.5 (mean 3.2, SD 1.12). Participants explained that they rated the strength of their habits as lower when they had not performed their physical activity that week. Nevertheless, the rising trend supports the idea that habits were successfully formed.

The weekly mean for the identity scores indicate whether participants think the exercise is becoming more “typically them.” Again, a linear regression analysis was calculated to predict identity based on time (ie, week). A significant regression equation was found ($F_{1,7}=16.938, P=.003$) with an $R^2$ of 0.74. This result indicates that participants successfully formed new beliefs about themselves.

Participants mentioned that they struggled with the concept of starting with small activities. Participants who successfully adhered for a longer period of time did change their attitudes toward MoveDaily over time; participants developed new experiences with MoveDaily. For example, 7 participants out of 26 (27%) who indicated during the initial interviews that they liked the concept, yet missed the competitive aspect, eventually dropped out. However, 6 other participants out of 26 (23%), who similarly indicated missing the competitive aspect, finished the 10-week period due to their interest in developing automaticity in physical activity. When continuing,
these participants developed similar attitudes as those who were positive about the concept from the start.

The interview data also indicated that participants became more aware of how they could integrate small activities into their daily lives. For example, whereas one of the participants used to not think about taking the stairs, elevator, or escalator, often resulting in not taking the stairs, she now very consciously took the stairs. This awareness might come from the focus on a certain identity, that is “I am a stairs person.” In other cases, participants seemed to become more aware of their health in general, similar to what occurred with the diary cards. One participant explained about his recent contact with a physiotherapist, another participant became a bit more aware of his diet, and sometimes participants even picked up sports (e.g., tennis) during the period of the study.

**Phase 3: Conceptual Design**

Data gathered during the user evaluation of the conceptual design of MoveDaily indicated a general positive reaction to the app. During observation, participants showed the ability to go through the different interaction scenarios and complete the required tasks with relative ease. While interacting with the prototype, all 4 participants (100%) indicated an interest in learning more about the habit-forming aspect of the product and appreciated the alternative approach of starting with small activities (see Table 1). The two most-pressed issues for improvement did not directly relate to the introduced identity approach itself. Participants indicated that they would recommend to (1) improve the user interface of the home page and (2) improve the copywriting within the app. Overall, participants seemed pleased with the usability of the high-fidelity prototype and with the identity-oriented concept.
Table 1. List of core themes and relative example quotes derived from qualitative analysis of the three phases.

<table>
<thead>
<tr>
<th>Phases, theme groups, and themes</th>
<th>Example quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 1</strong></td>
<td></td>
</tr>
<tr>
<td>Habit forming</td>
<td></td>
</tr>
<tr>
<td>Starting attitude toward starting small</td>
<td>“Four squats mean absolutely nothing!”</td>
</tr>
<tr>
<td>Conscious of the daily activity</td>
<td>“Seeing the value in starting small and how this can be built up, is quite motivating.”</td>
</tr>
<tr>
<td>Awareness of habit construct</td>
<td>“By changing my routine, I was capable of doing my exercise successfully more often. Before I had struggles with time and clients dropping by.”</td>
</tr>
<tr>
<td>Identification</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Other</td>
<td>N/A</td>
</tr>
<tr>
<td>Importance of a fun factor</td>
<td>“I do not really have an opinion about the stickers, but it is for sure more fun than simply writing.”</td>
</tr>
<tr>
<td></td>
<td>“It feels like the cards attempt to put an application’s functionalities within paper. I would prefer digital reminders, now I have to be aware of the piece of paper.”</td>
</tr>
<tr>
<td><strong>Phase 2</strong></td>
<td></td>
</tr>
<tr>
<td>Habit forming</td>
<td></td>
</tr>
<tr>
<td>Starting attitude toward starting small</td>
<td>“Sometimes I forgot whether or not I did my habit. If it would be more difficult, I would feel like it has more effect.”</td>
</tr>
<tr>
<td>Ease of consistency</td>
<td>“I think I often want to move but am unable to find the motivation. So in this sense, the concept speaks to me.”</td>
</tr>
<tr>
<td></td>
<td>“By doing my habit every day, I think I am strengthening the value of the thought &quot;Yes, I did it again today. I am good!!””</td>
</tr>
<tr>
<td>Feeling free of any pressure or performance</td>
<td>“It is nice to do something daily for myself.”</td>
</tr>
<tr>
<td></td>
<td>“It is also good for the average person; no gym talk like ‘You are a loser if you do not look like me.’ It guides you while providing enough freedom for you to create your own habit.”</td>
</tr>
<tr>
<td>Identification</td>
<td></td>
</tr>
<tr>
<td>Strong self-identity</td>
<td>“I have not really started to move more, but I have become more aware of the things I am doing. Like taking the stairs at the station is completely engrained in my behavior. Even when I am carrying a heavy bag during vacation, I still do not want to give up on taking the stairs.”</td>
</tr>
<tr>
<td>Adoption of a behavior</td>
<td>“I currently do not just take the stairs at work. I also take them at the station or wherever I am.”</td>
</tr>
<tr>
<td>Support in rewarding</td>
<td>“I fill in the form quite unconsciously. I am not sure if I really need it anymore, but I always do like the reminder to give myself a small compliment.”</td>
</tr>
<tr>
<td><strong>Phase 3</strong></td>
<td></td>
</tr>
<tr>
<td>Habit forming</td>
<td></td>
</tr>
<tr>
<td>Interest in learning</td>
<td>“I would be interested in more habit-forming tips, for example, how to create better triggers.”</td>
</tr>
<tr>
<td>Feeling free of any pressure or performance</td>
<td>“I appreciated the in-app feedback and how it explained that even when I failed, the behavior can be reshaped to fit me better.”</td>
</tr>
<tr>
<td>Identification</td>
<td></td>
</tr>
<tr>
<td>Difficulties with technical phrasing</td>
<td>“‘Developing a new identity’ sounds a bit heavy to me.”</td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
<tr>
<td>Engagement with gamification</td>
<td>“I like how you explore the world further, the more successful movements you make.”</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.
Discussion

Principal Findings
The first RtD phase (ie, diary cards) created a better understanding on how to translate the combination of the habit loop with the identity loop into a fitting and valuable intervention design. It also resulted in a confirmation that participants would prefer, and maybe even place more trust in, a digital intervention. In the second RtD phase, we found that high habit scores are indeed, as hypothesized, indicative of physical activity frequency. Similarly to forming habits in general, frequency is, thus, also important for forming strong identity-oriented habits. The third phase of this study, the conceptual design phase, showcases how an app that makes people more identity oriented could materialize. It is offered in the format of a functional game supporting people to become more physically active.

Taken together, the three RtD phases effectively show how theory on behavior change can be implemented in design by following an iterative approach. The presented RtD approach intends to show designers how the complex situation surrounding adherence to physical activity can be approached, how it can be framed and reframed, and how prototypes that address it can be iteratively developed. In this manner, the design approach allows for designers to learn what a design interpretation of a theoretical model could look like and what responses of future users might be. Thereon, the design could more easily be improved by a future design (ie, iteration).

This study followed an explorative and iterative approach. The number of participants was rather low and we mostly collected data based on self-report. Moreover, our participants were young and we did not measure or ask our participants about their fitness levels or amounts of daily or weekly exercise before participation. While this will have influenced the accuracy of the data on physical activity and limits extrapolation to a general population, our data provided rich insights about participants’ attitudes toward using the intervention and changing their behavior. In order to evaluate the long-term efficacy and influence of the inward orientation (ie, identity), a longitudinal study should be conducted that includes a more heterogeneous population and that collects actual log data from the mobile app and combines this with self-reports. Baretta et al [51] showed that such a longitudinal study can provide rich insights into how people experience certain app features and what they think is most beneficial to them over the longer term. Future research may also be aimed at determining what small exercises are most effective in changing people’s beliefs about themselves.

Conclusions
It is clear that design can influence human behavior, yet the understanding of designing for behavior change is still fragmented [52]. This paper provides deeper insights into designing for electronic health (eHealth) interventions and services and suggests a new way to shape motivation through the design of an intervention and adherence to physical activity. It introduces a new model to address the issue of behavior change in relation to physical activity and demonstrates how this model can be used in an iterative design approach that focuses on the development of a mobile health (mHealth) app that supports being more physically active. In this way, the described process serves as an example of a holistic approach to design for behavior change.

To the best of our knowledge, this is the first iterative design study in which the effects of adopting an identity approach to both motivation and physical activity were included and observed. Through the presented approach, which combines the habit loop with the identity loop, we presented initial promising results toward understanding how sustainable behavior change can be achieved. Habits serve as an important base for behavior change by having someone identify with the personal importance of a behavior repetitively. Furthermore, we demonstrated how such a theoretical idea can be explored in a three-step iterative RtD approach.

Authors’ Contributions
FH and GDSL discussed the setup of the RtD process throughout the research. FH was responsible for the study design and data collection and prepared a first draft of the manuscript. FH and GDSL worked on finalizing the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Example of a daily interaction with the web-based prototype.
[PDF File (Adobe PDF File), 3665 KB - mhealth_v8i6e16720_app1.pdf ]

Multimedia Appendix 2
Self-Report Habit Index (SRHI) questionnaire.
[PNG File, 182 KB - mhealth_v8i6e16720_app2.png ]

References


**Abbreviations**

- eHealth: electronic health
- mHealth: mobile health
- RtD: Research through Design
- SRHI: Self-Report Habit Index
- SRHI-P: Self-Report Habit Index-physical activity

©Floris Hooglugt, Geke D S Ludden. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 15.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Scalable System for Passively Monitoring Oral Health Behaviors Using Electronic Toothbrushes in the Home Setting: Development and Feasibility Study

Vivek Shetty1, DDS, DrMedDent; Douglas Morrison1,2, BA, MSc; Thomas Belin1,2, PhD; Timothy Hnat3, PhD; Santosh Kumar3, PhD

1Section of Oral & Maxillofacial Surgery, Department of Biomedical Engineering, University of California, Los Angeles, Los Angeles, CA, United States
2Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, United States
3Department of Computer Science, The University of Memphis, Memphis, TN, United States

Corresponding Author:
Vivek Shetty, DDS, DrMedDent
Section of Oral & Maxillofacial Surgery
Department of Biomedical Engineering
University of California, Los Angeles
Los Angeles, CA
United States
Phone: 1 3108255170
Email: vshetty@ucla.edu

Abstract

Background: Dental disease (including dental caries and periodontal disease) is largely preventable and closely linked to inadequate oral health behaviors. Digital health technologies have great potential for unobtrusively monitoring brushing behaviors in home settings and promoting optimal oral self-care routines at scale.

Objective: The aim of this study is to leverage the ubiquity of electronic toothbrushes and smartphones with the development of a Remote Oral Behaviors Assessment System (ROBAS) and evaluate its feasibility for passively tracking brushing behaviors in real-world settings.

Methods: We developed ROBAS by linking inertial sensors contained within consumer electronic toothbrushes to a scalable software platform comprised of a smartphone app linked to a cloud platform. First, the criterion validity of ROBAS for accurately capturing brushing details was established in a laboratory setting. Next, real-world performance and usability were evaluated in a stratified community sample of 32 participants who used ROBAS daily for 1 month and maintained a diary of their brushing episodes. Semistructured interviews at baseline and exit captured the user experience. We used regression models and Bland-Altman analyses to assess the criterion validity, functionality, accuracy, and consistency of ROBAS.

Results: Using a stopwatch as the criterion reference, ROBAS showed a mean absolute percent error (MAPE) of 1.8%, an estimated bias of 0.64 seconds that was not statistically distinguishable from zero (95% CI –0.93 to 2.22 seconds, SE 0.79), and a connection failure rate of 6.7% (95% CI 0.8%-22.1%, SE 4.6%). In real-world testing, ROBAS showed close agreement with the daily diary recordings of brushing episodes; estimated average discrepancies between the diary and ROBAS were 0.13 sessions per day (95% CI 0.01-0.26, SE 0.06), 8.0 seconds per brushing session (95% CI 1.4-14.7, SE 3.3), and 30 seconds of brushing per day (95% CI 9.3-50.1, SE 10.0). Retrospective self-reports produced substantially higher estimates of brushing frequency and duration compared to ROBAS measurements. Participants reported ROBAS was easy to use and expressed an interest in receiving ROBAS-delivered feedback on their brushing behaviors. Most participants were bothered by the use of an additional study phone, and some reported connectivity-related issues.

Conclusions: ROBAS has a high criterion validity for measuring oral health behaviors. It can accurately and reliably monitor brushing patterns in home settings for extended periods. Unobtrusive data collection through ROBAS sets the stage for automated coaching and optimization of oral self-care practices at the individual and population level.

(JMIR Mhealth Uhealth 2020;8(6):e17347) doi:10.2196/17347
KEYWORDS
health behaviors; oral self-care; digital tools; remote monitoring; passive measurement

Introduction

Although largely preventable, dental disease (including caries and periodontal disease) is extremely common and exacts a substantial personal and societal toll [1,2]. Dental disease is closely linked to poor oral hygiene behaviors (OHB). Considerable scientific evidence indicates that systematic, twice-a-day tooth brushing with a fluoridated toothpaste prevents accumulation of dental plaque (a sticky film containing bacteria) that leads to tooth decay, gum disease, and eventually, tooth loss [3-5]. However, this basic health behavior is not as widely and fully practiced as dentists and health organizations would like it to be [6]. Large population surveys reveal that a majority of individuals (45% to 67%) brush less frequently than twice a day, with brushing habits worsening with advancing age [7,8].

The essential predicament of traditional dental care is that what happens during the roughly 363 days of the year that typical patients spend outside of the dental clinic is of far greater consequence than the 2 days when they have clinic visits [9]. Self-reports rarely provide the care provider with an accurate picture of actual brushing behaviors because of distorted recall of what, for the vast majority of patients, is a low-salience activity as well as associated social desirability biases [10,11]. Furthermore, oral hygiene instructions delivered during dental care visits are not scalable and do not reliably inspire lasting habits because they are very generic; sporadic; and largely disconnected from the patient’s values, needs, and preferences. Ultimately, good oral health depends on the individual’s ability and willingness to carry out the mundane self-care behaviors at home. To achieve meaningful improvements across large, diverse populations, it is essential for providers to identify and support optimal oral health behaviors in the home setting.

The convergence of technological advances, deep penetration of digital devices, and a generational shift in how the technology is used and consumed provides unique opportunities to connect remotely and to engage with patients at a personal level [12,13]. Self-care technologies, including devices whose embedded sensors and analytic algorithms can track, analyze, and guide the user’s behaviors, are increasingly used to help individuals recognize and improve daily lifestyle choices that add up to affect their health [14-16]. Wearable devices like the Fitbit (Fitbit Inc) and the Apple Watch (Apple Inc) are prominent examples of these self-tracking technologies. Initially developed to help users measure daily activity, they have evolved into connected health platforms that seek to facilitate healthy behaviors by providing individuals with relevant feedback, timely and personalized cues, and motivational rewards, all of which may support health behavior change.

Building on technology’s potential to cultivate interest and awareness in mundane self-care behaviors, we set about creating a low-cost digital platform that could measure and track oral health behaviors in the lived environment and set the stage for automated, personalized coaching at population scale. Such a Remote Oral Behaviors Assessment System (ROBAS) would leverage the ubiquity of electronic toothbrushes and smartphones as well as sociotechnological trends in how digital devices are used. Our objective was to develop a ROBAS and to evaluate its performance and feasibility for passive tracking of brushing behaviors in real-world settings.

Methods

Development of ROBAS

ROBAS builds on a broadly available consumer-grade electronic toothbrush (Oral-B 7000; Procter & Gamble) as the data source for brushing behaviors (timing, duration, pressure applied). With the permission of the manufacturer, the application programming interface of the Oral-B toothbrush was adapted to expose the brushing data captured by the embedded accelerometer and pressure sensor. Collected data is transmitted over BLE (Bluetooth Low Energy) to a paired Android mobile device running the companion mCerebrum data collection app [17]. Developed by the Mobile Data to Knowledge (MD2K) Center, mCerebrum is a highly extensible, open-source platform that permits concomitant data collection from multiple sensors with real-time assessment of data quality [18]. Collected data is then uploaded to the secure Cerebral Cortex cloud [19] for remote monitoring of data yields and analytics. Visualization of time series data streams of brushing episodes and monitoring of sensor function and participant compliance is accomplished through an adaptation of the open platform Grafana dashboard (Figure 1) [20].
Research Design

The study protocol was reviewed and approved by the institutional review board, and all participants provided prior written informed consent. The first step was to verify the reliability and concurrent validity of ROBAS by comparing brushing data (duration) captured by ROBAS to data from a reference gold standard generated using a stopwatch. To assess the possibility of unacceptably poor calibration, 3 volunteer participants used ROBAS for 10 brushing sessions conducted over several days in a controlled laboratory setting. Each participant performed brushing sessions with ROBAS and a conventional stopwatch was used to record the start and stop time of each session.

Subsequently, the real-world performance and usability of ROBAS was evaluated in a community sample of 36 participants, stratified by gender, age group (18-29, 30-44, and ≥45 years), and self-reported tech-savviness (less tech-savvy or more tech-savvy). The original design contemplated balanced assignment of 3 participants to each of the 2x2x3 strata defined by gender, tech savviness and age group, with participants who dropped out early or who failed to conform to the study protocol replaced by participants from the same baseline-characteristic stratum.

In total, 2 participants dropped out at a late phase of the study and an additional 2 participants completed their exit interviews but failed to return their diaries; this left a final sample of 32 participants with 1 month of ROBAS data and diary recordings (Table 1). The balanced stratified sampling design allowed estimation of main effects as well as two-way interaction effects for the stratum-defining characteristics with a relatively modest sample size.

Table 1. Participants by recruitment stratum.

<table>
<thead>
<tr>
<th>Age (years) and gender</th>
<th>Less tech-savvy</th>
<th>More tech-savvy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>18-29</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
<td>2a</td>
</tr>
<tr>
<td><strong>30-44</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>2b</td>
<td>3</td>
</tr>
<tr>
<td><strong>≥45</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>2b</td>
<td>2a</td>
</tr>
</tbody>
</table>

a A participant completed the exit interview but did not return their brushing diary.

b A participant dropped out at a late stage.
At the baseline visit, participants were queried about their brushing habits including the frequency, duration, and time of day. Next, they were provided a daily diary and an electronic toothbrush paired with a dedicated Android phone (Samsung Galaxy 6). Participants were coached in the use of ROBAS by the study staff and received clarifications on its features. Participants were instructed to use ROBAS exclusively for 4 weeks at home and to time their brushing details (start time, end time, and duration) using a stopwatch and record these details in the diary.

At the exit visit, participants were asked to retrospectively estimate their brushing frequency. A brief, semistructured qualitative interview was used to assess participants’ general reactions to ROBAS, generate feedback on their experiences with the mCerebrum app, and learn about future interests in phone-based reporting of oral health behaviors. In addition to detailed observation notes by the interviewer, the interviews were recorded, anonymized and transcribed for subsequent thematic analysis. As part of the interview, participants were asked 10 multiple-choice questions with a 5-point Likert scale (Strongly agree, Agree, Neither agree nor disagree, Disagree, Strongly disagree) as response options. For easier interpretation, these answers were subsequently recoded into a favorability scale (Strongly Favorable, Favorable, Neutral, Unfavorable, Strongly Unfavorable).

**Data Analysis**

All quantitative analyses were conducted using the R software system (R Foundation for Statistical Computing) [21]. We used regression models and Bland-Altman analyses [22] to assess the criterion validity, functionality, accuracy, and consistency of ROBAS. For criterion validation, discrepancies in per-session brushing duration (as recorded with ROBAS versus a stopwatch) were modeled using linear regression, with random intercepts for session nested in participant and a fixed effect for data source. For feasibility testing, brushing data (recorded with ROBAS and daily diary) were compared in terms of 3 participant-level outcomes: number of sessions recorded per day, mean per-session minutes of brushing, and mean daily minutes of brushing. Scatterplots and Bland-Altman plots were used to evaluate agreement between self-reported and measured data. The rates of reported ROBAS-related connectivity issues were also extracted from the participant diaries.

The average frequency of brushing per day was compared using values from ROBAS, diaries and self-reports provided at baseline and exit interviews. Average per-session duration of brushing was compared between ROBAS and baseline (duration was not surveyed at exit).

Per-participant scatterplots were used to explore patterns of diary recordings. The average discrepancy in per-session brushing duration among the matched sessions was estimated using a random effects model to account for repeated measurements grouped by participant.

Usability surveys were summarized across all participants, and the relationships between baseline characteristics and usability ratings were explored using univariate and multivariate linear regression models, treating ratings of Strongly Favorable as +2, Favorable as +1, Neither Favorable nor Unfavorable as 0, Unfavorable as −1, and Strongly Unfavorable as −2.

**Results**

**Laboratory-Based Criterion**

Each of the 3 participants completed 10 brushing sessions. In two instances, the brush failed to connect with the phone when activated, resulting in a connection failure rate of 6.7% (95% CI 0.8%–22.1%, SE 4.6%). There were thus 28 brushing sessions with data usable for analysis. Using the stopwatch recordings as criterion reference, ROBAS showed a mean absolute percent error (MAPE) of 1.8% and an estimated bias of 0.64 seconds that was not statistically distinguishable from zero; the data were not compatible with a hypothesis of a large magnitude of bias (95% CI –0.93 to 2.22 seconds, SE 0.79 seconds). Brushing durations captured by the stopwatch averaged 117.9 seconds compared to an average duration of 118.5 seconds as recorded by ROBAS. The estimated root mean squared error was 4.2 seconds.

**Performance & Feasibility Testing**

**Participant Characteristics**

A total of 32 participants completed the study and provided ROBAS data as well as diary recordings of their brushing episodes (Table 1). None of the participants had previous experience with an electronic toothbrush. All reported smartphone ownership; about half the participants used their phones extensively for internet browsing, banking, and social media activities (more tech-savvy).

**Feasibility Testing**

In the home setting, ROBAS recorded 1242 brushing sessions (38.81 brushing sessions/participant) in contrast to the 1362 sessions (42.56 brushing sessions/participant) recorded in the participant diaries; a mean discrepancy of 3.75 sessions per participant (95% CI –0.93 to 2.22 seconds, SE 0.79 seconds). In total, 1095 sessions were recorded in both ROBAS and the diary, 147 sessions were recorded in ROBAS only, and 267 sessions were recorded in the diaries only (Table 2).
Table 2. Brushing session counts according to ROBAS and diaries.

<table>
<thead>
<tr>
<th>Diary</th>
<th>ROBAS Recorded</th>
<th>Not recorded</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recorded</td>
<td>1095</td>
<td>267</td>
<td>1362</td>
</tr>
<tr>
<td>Not recorded</td>
<td>147</td>
<td>0</td>
<td>147</td>
</tr>
<tr>
<td>Total</td>
<td>1242</td>
<td>267</td>
<td>1509</td>
</tr>
</tbody>
</table>

For the most part, the diary records of the sessions corresponded closely to that captured by ROBAS (Figure 2). However, a few participants recorded substantially more brushing sessions in the diary than in ROBAS. The Bland-Altman method was used to examine the limits of agreement between the data on brushing sessions from ROBAS and the actual sessions recorded in the diaries. The plot of this data showed an apparent relationship between the average of the two measurements and the distribution of the discrepancies; it appeared that the variance of the discrepancies was larger for participants whose measurements averaged 30 to 40 sessions, compared with participants with higher or lower averages (Figure 3). The source of this heteroscedasticity is not immediately obvious. Some of the discrepancies appeared related to technical issues; participants logged 85 instances of connectivity issues between the brush and the phone app (average 6.83%). However, the relationship between technical issues and discrepancies was not entirely clear; some of the participants with the largest discrepancies reported many technical issues, but others did not report technical issues at all (Multimedia Appendix 1).

Figure 2. Brushing sessions recorded in participant diaries and by ROBAS. Dashed line denotes y=x. ROBAS: Remote Oral Behaviors Assessment System.
At both baseline and exit interviews, most participants self-reported at least two brushing sessions per day. However, data captured by ROBAS data revealed that few participants actually averaged two sessions per day. Retrospective self-reports of brushing sessions/day at the exit interview were substantially higher than the ROBAS and diary data (Figures 4 and 5). On average, participants recollected at least 2.0 sessions/day (95% CI 1.8-2.1, SE 0.091). However, ROBAS documented a mean of 1.4 sessions/day (95% CI 1.2-1.6, SE 0.092) and the diaries showed a mean of 1.5 sessions/day (95% CI 1.4-1.7, SE 0.082). The estimated mean discrepancy between retrospective self-report and ROBAS was 0.57 sessions/day (95% CI 0.35-0.79, SE 0.11), and the estimated mean discrepancy between retrospective self-report and the diary was 0.43 sessions/day (95% CI 0.25-0.62, SE 0.09).
Figure 4. Mean daily brushing sessions by participants, as measured by ROBAS versus retrospective self-report. ROBAS: Remote Oral Behaviors Assessment System.

Figure 5. Mean daily brushing sessions, as measured by brushing diaries versus retrospective self-report.
Brushing Duration Per Session

The mean discrepancy between ROBAS and diary recordings of brushing duration/session was 8.0 seconds (95% CI 1.4-14.7, SE 3.3); ROBAS recorded an average per-session brushing duration of 137.41 seconds (95% CI 123.37-151.46, SE 6.89), whereas the diaries showed a duration 145.44 seconds (95% CI 131.12-159.76, SE 7.02). This discrepancy may be artificially small because of the unintended display of the recorded brushing duration on the ROBAS user interface.

Individual Brushing Sessions

The mean daily brushing duration recorded by ROBAS was 3.14 minutes (95% CI 2.65-3.63, SE 0.239), whereas the mean daily brushing duration recorded in the diaries was 3.64 minutes (95% CI 3.15-4.12, SE 0.237). The mean discrepancy between the diary and ROBAS was thus 30 seconds (95% CI 9.3-50.1, SE 10.0). Most of the individual brushing sessions recorded by ROBAS and the diaries were in close agreement (Figure 6). Among the matched sessions, the estimated average discrepancy in session duration between the diary and ROBAS was 8.6 seconds (95% CI 2.3-15.0, SE 3.2); 39% of the matched records had exactly the same elapsed time recorded in both ROBAS and the diary. Again, this close concordance may be due to the user interface design, in which the ROBAS app showed the time elapsed.

Figure 6. Durations of individual brushing sessions, as measured by ROBAS and brushing diaries, grouped by participant. Horizontal alignment of some data points suggests guessed approximations. ROBAS: Remote Oral Behaviors Assessment System.

Several diaries showed evidence of approximated or rounded records suggesting that the entries were estimated much later than the actual event; 4 participants recorded the same brushing duration (eg, 2 minutes) for most (>90%) of their brushing sessions, indicating that the diary entries were not made contemporaneously. There was no apparent relationship between the tendency to report the same brushing duration and the frequency of reporting glitches (Figure 7).
**Usability Surveys**

From a quantitative standpoint, satisfaction with ROBAS was uniformly high (Figure 8). Most participants (85%) found the system relatively easy to use and enjoyed the experience of the electronic toothbrush. A majority expressed an interest in receiving feedback about their brushing behaviors (91%) and indicated that they would recommend ROBAS to a family member or friend (86%). When the exit interview responses were stratified by each of the baseline characteristics, none of the questions showed evidence of differences in mean by gender, age group, or tech-savviness ($\alpha=.05$). However, a sizeable segment (68%) reported that they found the additional dedicated study phone to be burdensome. Some participants (6.8%) reported technical issues involving connectivity (eg, participant was unable to connect the brush to the phone app, lag in connection) or were bothered by a short battery life.
Figure 8. Exit interview usability survey responses; the percentages of negative, neutral, and positive responses are printed on the left, center, and right, respectively, of each color bar.

Discussion

Overview

Our goal was to develop and evaluate a system (ROBAS) that would help expand oral care beyond the confines of the dental clinic. Our study showed that ROBAS is accurate and reliable in its ability to passively capture oral self-care practices (when, how long, sessions per day) in the home setting. Specifically, ROBAS (1) had very high criterion validity and demonstrated close agreement with stopwatch measurements of brushing sessions; (2) was less burdensome and more reliable than daily diaries; (3) provided a more objective and accurate representation of brushing behaviors than retrospective end-of-study recall; and (4) was generally well accepted by its users.

In a controlled laboratory setting, the details of the brushing sessions acquired by ROBAS closely approximated the ground truth values captured by the stopwatch. The high overall aggregated accuracy (<2% MAPE) and the good repeatability across brushing sessions showed that the automated data collected by ROBAS was equivalent to the manual stopwatch recordings. The concordance between the brushing session details captured by ROBAS and the daily diary records largely held up in the home setting. However, the per-session and per-day brushing durations estimated by the diaries and ROBAS showed modest discrepancies. The inconsistencies likely arose from two sources: (1) differences due to reporting errors, and (2) differences arising from technical issues.

The patterns of overreporting manifest in the diary logs of a few participants suggested a degree of fabrication with several diary entries revealing approximations of brushing duration. Some of the participants recorded “2 minutes” for every brushing session, indicating that the diary entries were not made immediately after the event. These findings underscore the challenges of depending on the willingness of participants to meticulously record details of fixed events for extended periods of time. Diary data can be of questionable reliability because they are burdensome to gather, often incomplete because of illegibility or loss, and susceptible to invention [23]. Similarly, the retrospective recall of brushing sessions/day at the exit interviews were substantially higher than that captured by ROBAS and the diary. Our results highlight the unreliability of self-reports when it comes to the recall of health behaviors of low salience [24-26]. Inaccurate estimates, originating from recall bias or social desirability bias, can lead to underestimation of risk parameters during routine dental care visits and result in erroneous self-care recommendations [27].

Technical issues related to the ROBAS prototype also contributed to an underrecording of brushing events. The glitch rate (approximately 6.8%) encountered could be related to a number of factors including failure of the toothbrush to connect to the mCerebrum app, excessive battery drain, or the participant forgetting to charge the study phone. Connectivity may have suffered from the fact that the proprietary bridging software provided by the manufacturer (Oral-B/Procter & Gamble) was written by a third-party vendor and prevented the mCerebrum app from fully controlling the brush connection process and hindered the identification and correction of root causes of connectivity failures. As our experience showed, integrating software for Android phones to communicate with a specific device’s Bluetooth stack at a low level is a process fraught with frustration as developers can implement their own custom Bluetooth interface. Resetting the devices usually reestablished connectivity; however, acceptance of any self-care technology...
would drop sharply if the end user is required to take frequent corrective actions [28]. Iterative versions of ROBAS have improved the connectivity issue. Additionally, low battery levels now trigger alerts to remind users to proactively charge the device’s batteries.

From a usability and acceptance standpoint, participants were generally very satisfied with ROBAS and enjoyed the new experience of an electronic toothbrush. Most expressed an interest in receiving brushing summaries along with personalized actionable suggestions. Interestingly, the participants did not express any concerns about privacy issues. If anything, participants had a very relaxed attitude toward reporting and having their activities recorded in our specific setting. Many participants disliked the requirement of an additional study phone. A subset reported frustration with the sporadic technical issues (eg, unable to connect the brush to the phone app, lag in establishing connection) and short battery life of the study phone. Comments to the open-ended exit interviews revealed that participants largely preferred the automaticity of ROBAS to the tedium of maintaining a daily diary log of their brushing sessions.

Our design objective was to develop a human activity recognition system that would weave into the fabric of everyday life to extend the reach and continuity of health care. To that end, we leveraged everyday devices (electronic toothbrushes and smartphones) and habitual behaviors (brushing) to passively monitor oral self-care practices in the lived environment. Tools, like ROBAS, that introduce passive measurement into the delivery of dental care have considerable potential in allowing actionable insights on actual oral self-care practices in the home setting. The objective data gathered by the sensing technology could eventually drive computationally driven, adaptive behavioral nudges that automatically adjust to the individuals’ changing behavior, history, and environmental contexts [29]. Temporal changes in the data sets would help determine who engages in the interventions, how they engage, and factors that promote engagement. Furthermore, patterns around self-care behaviors would allow care providers and health systems to proactively identify and focus on those most at risk to ultimately mitigate the high burden and costs of dental disease.

We feel that ROBAS is the closest alternative to direct observation of brushing behaviors and a valuable new tool for researchers interested in investigating or measuring oral self-care behavior. Ultimately, the usability of the system will determine user adoption. Based on our study results, several technological improvements to the ROBAS prototype have been carried out. These include an updated smart toothbrush Bluetooth software development kit (SDK), the addition of a gyroscope within the smart toothbrush, and more robust mechanisms to handle and detect the Bluetooth connection challenges on Android smartphones. The use of a dedicated study phone, a perceived burden in the ROBAS testing, has been abandoned in favor of a Bring Your Own Device (BYOD) model in which participants install the mCerebrum app on their own mobile phone. Although this introduces a technical complexity in that participants are required to provide a phone that is fit for purpose, the pragmatic move greatly increases the pool of participants available for follow-up studies and decreases the provisioning costs (ie, costs of a dedicated phone) and supply and training issues (ie, delivering the devices to the participants and training them).

The promising results notwithstanding, our study had some limitations. As discussed above, some participants may have used the brushing durations displayed by the ROBAS app instead of independently measuring brushing duration using a separate stopwatch, artificially increasing the concordance between the diaries and the ROBAS data. Additionally, selection bias influences the levels of measurement agreement found in this study. Although we recruited from the community, study participants had to opt into this study, attend an orientation visit, complete the study protocol, and return for an exit interview. This requirement may have led to a sample with higher levels of success in the use of ROBAS and accuracy in recording the diaries than the general population.

**Conclusion**

Based on our study findings, ROBAS has a high criterion validity for measuring oral health behaviors. It can accurately and reliably monitor brushing behaviors in the home setting for extended periods. Unobtrusive data collection through ROBAS sets the stage for automated coaching and optimization of oral self-care practices for each individual across the population.

**Acknowledgments**

Research reported in this publication was supported by the National Institute of Dental and Craniofacial Research of the National Institutes of Health under award number R01DE025244. The authors wish to acknowledge the material and technical support provided by Mr Ingo Vetter and Dr Leslie Winston from Oral-B/Procter & Gamble. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Procter & Gamble.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Discrepancies in number of sessions recorded (diary – ROBAS) versus number of glitches reported in diary. ROBAS: Remote Oral Behaviors Assessment System.

[ PNG File , 19 KB - mhealth_v86f17347_app1.png ]
References


14. Shetty et alJMIR MHEALTH AND UHEALTH

15. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. mHealth 2017 Nov 27;3:51-51 [FREE Full text] [doi: 10.3390/mhealth.2017.08.06] [Medline: 29430455]


18. MD2K Center of Excellence for Mobile Sensor Data-to-Knowledge. URL: https://md2k.org/software.html [accessed 2020-05-14]


Abbreviations

- **BLE**: Bluetooth Low Energy
- **BYOD**: Bring Your Own Device
- **MAPE**: mean absolute percent error
- **OHB**: oral hygiene behavior(s)
- **ROBAS**: Remote Oral Behaviors Assessment System
- **SDK**: software development kit
Hospital Bring-Your-Own-Device Security Challenges and Solutions: Systematic Review of Gray Literature

Tafheem Ahmad Wani¹, MIT; Antonette Mendoza¹, PhD; Kathleen Gray², PhD

¹School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
²Centre for Digital Transformation of Health, The University of Melbourne, Melbourne, Australia

Corresponding Author:
Tafheem Ahmad Wani, MIT
School of Computing and Information Systems
The University of Melbourne
Level 6, Doug McDonell Building
Parkville
Melbourne, 3010
Australia
Phone: 61 451906170
Email: twani@student.unimelb.edu.au

Abstract

Background: As familiarity with and convenience of using personal devices in hospitals help improve the productivity, efficiency, and workflow of hospital staff, the health care bring-your-own-device (BYOD) market is growing consistently. However, security concerns owing to the lack of control over the personal mobile devices of staff, which may contain sensitive data such as personal health information of patients, make it one of the biggest health care information technology (IT) challenges for hospital administrations.

Objective: Given that the hospital BYOD security has not been adequately addressed in peer-reviewed literature, the aim of this paper was to identify key security challenges associated with hospital BYOD usage as well as relevant solutions that can cater to the identified issues by reviewing gray literature. Therefore, this research will provide additional practical insights from current BYOD practices.

Methods: A comprehensive gray literature review was conducted, which followed the stepwise guidelines and quality assessment criteria set out by Garousi et al. The searched literature included tier 1 sources such as health care cybersecurity market reports, white papers, guidelines, policies, and frameworks as well as tier 2 sources such as credible and reputed health IT magazines, databases, and news articles. Moreover, a deductive thematic analysis was conducted to organize the findings based on Schlarman’s People Policy Technology model, promoting a holistic understanding of hospitals’ BYOD security issues and solutions.

Results: A total of 51 sources were found to match the designed eligibility criteria. From these studies, several sociotechnical issues were identified. The major challenges identified were the use of devices with insufficient security controls by hospital staff, lack of control or visibility for the management to maintain security requirements, lack of awareness among hospital staff, lack of direction or guidance for BYOD usage, poor user experience, maintenance of legal requirements, shortage of cybersecurity skills, and loss of devices. Although technologies such as mobile device management, unified endpoint management, containerization, and virtual private network allow better BYOD security management in hospitals, policies and people management measures such as strong security culture and staff awareness and training improve staff commitment in protecting hospital data.

Conclusions: The findings suggest that to optimize BYOD security management in hospitals, all 3 dimensions of the security process (people, policy, and technology) need to be given equal emphasis. As the nature of cybersecurity attacks is becoming more complex, all dimensions should work in close alignment with each other. This means that with the modernization of BYOD technology, BYOD strategy, governance, education, and relevant policies and procedures also need to adapt accordingly.

(JMIR Mhealth Uhealth 2020;8(6):e18175) doi:10.2196/18175

KEYWORDS
BYOD; bring-your-own-device; health care facilities; mhealth; mobile phone; confidentiality; computer security
**Introduction**

**Background**

Bring-your-own-device (BYOD) is a term that refers to the use of personal devices by employees for professional purposes. These devices typically include smartphones, tablets, and laptops and may even include internet of things (IoT) devices such as wearables. Health care is one of the leading industries driving BYOD usage [1-5]. Health care professionals are known to use personal mobile devices for work such as documenting clinical notes; accessing medical records, drug information, or test results; time-tabling; communicating with other staff as well as with patients; and looking up reference resources [1,6-9]. It is suggested that BYOD saves time and improves the productivity of clinicians, makes patient care more efficient through better care coordination and continuity, saves device procurement costs for health care organizations, and may even reduce hospital admission rates [10-12].

However, one of the key issues that BYOD faces is data security. The health care industry sees the greatest number of data breaches among all major industries around the world [13,14]. In part, health care is a target of cybercriminals for various reasons; for instance, medical credentials are said to be sold in the black market, especially the dark web, for over US $1000 [15,16].

One of the primary reasons for health care data breaches is BYOD itself. Hospitals may have little or no control over the security of their employees’ personal mobile devices, which may contain sensitive organizational data such as patient information. Hospitals also do not have any control over a user’s nonwork-related activity on their BYOD device, as ownership lies with the employee. In addition, health care IoT devices such as personal wearables are growing at an exponential rate, and with each device added to the hospital network, the chance of breach increases. Furthermore, given the highly regulated nature of the health care industry, which enforces strict measures to protect patient information, health care organizations face a heavy task of compliance with health data protection laws [17-19]. In short, BYOD security is “one of the biggest headaches for healthcare IT management” [20].

**Objectives**

There has been little research into BYOD security, especially in health care [21-23]. Our previous literature review of hospital BYOD security issues and risk mitigation found mostly expert commentaries and a dearth of real-life studies in clinical settings [17]. A limitation of our previous review was that only peer-reviewed literature was considered.

Therefore, the aim of this study was to investigate the challenges and solutions of hospital BYOD security by reviewing the gray literature. The aim of this paper meets the criteria set out by Garousi et al [24] for including gray literature in research: to reduce the gap between academic research and industry practices, to provide perspectives missing from peer-reviewed research, and to provide practical insights about hospitals’ BYOD usage.

**Methods**

This gray literature review followed the stepwise guidelines set out by Garousi et al [24].

**Search Process**

First, specialized and credible health information technology (IT) sources, which include magazines, databases, and news sources such as Xtelligent Healthcare Media (HITInfrastructure and HealthITSecurity), Pulse IT Communications, Healthcare IT News Australia, and Healthcare Information and Management Systems Society (HIMSS) media were searched.

Second, other tier 1 and tier 2 gray literature sources searched from Google and the market research platform Gartner, which fit the quality assessment criteria of Garousi et al [24], were considered. Tier 1 sources include highly credible sources where knowledge and authority of the source are well established and where content is produced in conformance with well-defined criteria [25]. These included mobile security white papers and reports, national health care department guidelines or policies, BYOD market research reports, and frameworks from reputable agencies and organizations [24]. Tier 2 sources include sources where knowledge and authority of the source are moderately credible [25]. These included news articles, company annual reports, and document presentations [24]. For these sources, health care terms were used in addition to the terms mentioned above, as part of the search string. These include terms such as “health,” “healthcare,” and “hospital.” Only the top 100 search results from tier 1 and 2 sources were inspected, as saturation of concepts was observed after this.

Finally, some articles were also extracted through snowballing of links or citations provided in the abovementioned sources.

**Quality Assessment and Eligibility Criteria**

Only articles that fit the quality assessment criteria established by Garousi et al [24] were chosen for the study. Articles were assessed based on the authority of source, method, date, objectivity, novelty, and impact to determine their suitability for this study. Articles from credible and reputable sources (first- and second-tier gray literature), with a clear objective and adding a unique perspective to the research or corroborating previous scientific evidence, were included.

Only gray literature articles published from 2016 to 2019 in the English language were considered eligible. This ensured contemporaneity and practical relevance of this research, given that BYOD security management has seen significant changes during this period, such as an increase in the number and type of data breaches as well as improvements in technology. In addition, this study was limited to mid- and large-size clinical settings, mainly hospitals and community health centers; smaller settings with budget or technological constraints to invest in BYOD security, such as private practices, were excluded. Finally, eligibility was strictly limited to security issues related to the usage of BYOD in hospitals. Issues such as bandwidth, availability, device interference, and medical infection risks were excluded.
Overall, 51 articles were included, as shown in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart [26] in Figure 1.

Figure 1. Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart.

Data Extraction and Synthesis
Deductive thematic analysis was conducted using the People Policy Technology (PPT) model by Schlarman [27] to organize the findings [28] in accordance with the previous review conducted by the authors [17]. The PPT model breaks down the security process into 3 core elements: people responsible for executing and supporting the security process; policy used for explaining supporting procedures and providing a clear direction for ideal security behavior; and technology, which includes products, tools, or materials used for supporting the security process [27]. Previous studies indicate that to optimize the cybersecurity controls, the alignment between technical and social dimensions is necessary. Therefore, a holistic approach is required to completely understand the security process [28-31]. Experts also say that the technocentric nature of the present cybersecurity practices has increased the success of cyberattacks as the social dimension of security is relatively neglected [32]. This makes the PPT model very useful as it advocates for equal emphasis on all 3 dimensions (people, policy, and technology) of the security process.

Results

Characteristics of the Included Gray Literature
Table 1 summarizes the types of articles included. The detailed characteristics of each article are provided in Multimedia Appendix 1.
Table 1. Characteristics of the included gray literature (N=51).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td>7 (14)</td>
</tr>
<tr>
<td>2018</td>
<td>12 (24)</td>
</tr>
<tr>
<td>2017</td>
<td>22 (43)</td>
</tr>
<tr>
<td>2016</td>
<td>10 (20)</td>
</tr>
<tr>
<td><strong>Type of study</strong></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>29 (57)</td>
</tr>
<tr>
<td>Secondary</td>
<td>22 (43)</td>
</tr>
<tr>
<td><strong>Outlet source</strong></td>
<td></td>
</tr>
<tr>
<td>Market report</td>
<td>12 (24)</td>
</tr>
<tr>
<td>News article</td>
<td>17 (33)</td>
</tr>
<tr>
<td>Opinion post</td>
<td>7 (14)</td>
</tr>
<tr>
<td>Legislation</td>
<td>6 (12)</td>
</tr>
<tr>
<td>White paper</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Policy document</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Newsletter post</td>
<td>2 (4)</td>
</tr>
<tr>
<td><strong>Outlet type</strong></td>
<td></td>
</tr>
<tr>
<td>First-tier gray literature</td>
<td>14 (27)</td>
</tr>
<tr>
<td>Second-tier gray literature</td>
<td>37 (73)</td>
</tr>
</tbody>
</table>

### Hospital Bring-Your-Own-Device Security Challenges

Hospitals face several data security challenges owing to the use of BYOD. These challenges are sociotechnical in nature. Therefore, based on the PPT model, this study classifies them as technology, policy, and people challenges [27].

#### Hospital Bring-Your-Own-Device Technology Challenges

The following sections discuss the key technical challenges associated with the use of BYOD in hospitals, which pose a threat to data security.

**Devices With Insufficient Security Control**

Personal mobile devices (BYOD devices) used by health care professionals lack the security controls and visibility of company-owned devices that may have preinstalled security settings and enable better security management [33,34]. For example, in a 2016 survey of US health care organizations, 11% of doctor-owned personal devices that stored patient data had highly vulnerable operating systems that were either outdated or jailbroken/rooted [35]. Although outdated versions may lack the necessary resistance against modern security attacks, jailbroken or rooted devices forcefully block security controls in lieu of additional functionality/control [34].

**Device Locking/Authentication**

According to the same survey, 14% of the devices owned by doctors contained some form of patient data, yet had no device locking or authentication mechanism for protecting sensitive information such as passwords, pattern locks, or biometric authentication [35].

#### App Security

The same survey also revealed that about 27.79 million mobile devices that had installed medical apps were infected with at least one high-risk malware, through downloading vulnerable apps from unregulated app stores [35].

Overall, 46% of the doctors shared patient data through picture messaging, 65% through SMS, and 33% via WhatsApp, according to the mentioned survey [35]. Similarly, 87% of the staff at a National Health Services (NHS)-based hospital in the United Kingdom were found to use such apps to discuss patient cases at work [36]. Consequently, their patients’ health information might be viewed by their colleagues or family members who have access to these platforms [37].

#### Network Security

The Skycurie survey also revealed that 39% of the devices used by doctors for their day-to-day practice were susceptible to network attacks by the fourth month of using the device [35], typically when the clinicians connected their devices to unsecure networks such as public hotspots [34,38].

### Hospital Bring-Your-Own-Device Policy Challenges

The following sections discuss the key BYOD policy-related challenges in hospitals.
Lack of Direction

About 62% of health care executives from US health care organizations stated that their hospitals either do not have a BYOD policy or that they are not aware of it [39]. Absence of a dedicated BYOD policy or BYOD program means that there may be a lack of clarity about associated issues such as corporate chain of responsibility, data ownership, data protection, prerequisites for device enrollment, access control, or clinical communication and compliance [33,40].

Legislative Compliance

Given the highly regulated nature of the health care industry, health care organizations need to maintain strict compliance with privacy laws. These laws are intended to protect patient privacy and provide transparency to patients in terms of who uses their data and how they are used or transmitted. They make it compulsory for health care organizations to strictly enforce strong and appropriate security controls and also define standard operating procedures. The notification of breaches to the government has also become compulsory under health data privacy laws [41]. Examples of such laws include the Australian Privacy Principles of 1988 and the Healthcare Identifiers Act of 2010 in Australia, the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States, the General Data Protection Regulation (GDPR) in the European Union, and the Personal Health Information Protection Act (PHIPA) and the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada [42-47].

Penalties

In case of failure to provide adequate security safeguards, or if found responsible for data breaches, health care organizations may face heavy penalties from the government, in addition to reputational damages. For instance, personal data of 3800 patients at a children’s hospital in Dallas, United States, were accessed from an international airport from an unencrypted, nonpassword-protected Blackberry device, which led to a fine of US $3.2 million over patient privacy breaches [48]. In another example, a lost laptop owned by an employee at a Pennsylvania-based cardiology center led to a breach of 1391 electronic patient records, for which the center was fined US $2.5 million [49].

Hospital Bring-Your-Own-Device People Challenges

The following sections discuss the key social or people-related challenges associated with the use of BYOD in hospitals.

Inappropriate Behavior

Health care is the only industry where insider threats such as human error and system misuse are more prevalent than external threats such as hacking. Overall, 35% of all insider threats in the health care industry are attributed to human error, whereas 24% of them occur because of system misuse [14]. BYOD is a major contributor to human error as it converges private and organizational data, thus increasing the chance of unintentionally sending patients’ information to the device owner’s friend or family member [37]. System misuse occurs when employees abuse the authority or permissions given to them, for example, retrieving personal information about a patient for purposes not related to health care [14]. BYOD limits the control of hospitals in managing sensitive organizational data and, therefore, increases the chances for system misuse to occur.

Lack of Awareness

Employee awareness is a critically important component of any BYOD program [14]. Phishing scams, fake tech support requests, and ransomware attacks have been successfully used in recent times [41,50]. In 2017, HIMSS Analytics revealed that 80% of the surveyed health IT executives rated employee awareness as the foremost concern related to health care data security [50,51].

Poor User Experience

If clinicians have to go through complex security procedures such as typing in long passwords or logging in repeatedly after periods of inactivity, they will use work-arounds instead, such as using common or easy-to-remember passwords, sharing passwords with colleagues, or using unauthorized/banned messaging apps such as WhatsApp for communication to minimize lost time, which is again a threat to BYOD security [38,52].

Skills Shortage

Overall, 82% of IT executives in a survey of 8 developed countries, including the United States, the United Kingdom, and Australia, said that there is a shortage of cybersecurity skills, and 76% of IT executives believed that their government was not investing sufficiently in cybersecurity talent. In addition, 25% of the respondents claimed that a lack of cybersecurity staff made their organization susceptible to cyberattacks [53].

Penalties

In case of failure to provide adequate security safeguards, or if found responsible for data breaches, health care organizations may face heavy penalties from the government, in addition to reputational damages. For instance, personal data of 3800 patients at a children’s hospital in Dallas, United States, were accessed from an international airport from an unencrypted, nonpassword-protected Blackberry device, which led to a fine of US $3.2 million over patient privacy breaches [48]. In another example, a lost laptop owned by an employee at a Pennsylvania-based cardiology center led to a breach of 1391 electronic patient records, for which the center was fined US $2.5 million [49].

Hospital Bring-Your-Own-Device Security Solutions

As discussed, to curb BYOD security risks, a holistic approach is required, with equal emphasis on technology, people, and policy-based measures. This section, therefore, classifies the solutions based on the PPT elements of the security process [27].

Hospital Bring-Your-Own-Device Technology Solutions

Technology is one of the core components of the security process, which can aid in efficiently managing BYOD security. The following are important technologies used to curb BYOD security risks.

Mobile Device Management

Mobile device management (MDM) is a platform used to manage devices existing within an enterprise centrally. It performs functions such as automation of device registration and deregistration as well as updating or patching of BYOD devices by remotely installing secure configurations, settings, and policies [34]. Moreover, MDM also automates remote installation of enterprise apps such as antivirus or antimalware over-the-air onto devices or scans enterprise networks for vulnerabilities [54,55]. Furthermore, it also automates the enforcement of organizational policies such as enabling screen lock or log-off functionalities; encrypting hospital data; securing remote connections through virtual private networks (VPNs); tracking device location; wiping, locking, and securing devices
remotely; and whitelisting and blacklisting apps and devices such as jailbroken/rooted devices or unapproved third-party apps [33,56,57].

**Containerization**

Containerization allows logical separation of organizational and personal data. This means that the hospital will only have control over the container where the hospital data reside, rather than the whole device. Containers are encrypted and therefore protect sensitive patient data that may reside on the employee’s device. The hospital can scan these data for viruses. It can even lock or delete the data remotely, while keeping the personal data intact [52]. In addition, containers also allow separate data backups for both personal and organizational data. Personal data can be uploaded on a personal cloud, whereas hospital data are uploaded on corporate or private cloud. IT administrators retain full control of the containers, and the need to manage the whole device is eliminated [52,58].

**Virtual Desktop Infrastructure**

Virtual desktop infrastructure eliminates BYOD security risks by removing the need to store hospital data on employees’ personal devices. It can provide access to hospital data through remote servers owned by the hospital, which can be connected to via a VPN after logging in remotely with proper credentials [34].

**Identity and Access Management**

Identity and access management (IAM) technologies ensure appropriate access to verified BYOD users through strong authentication, authorization, and access control mechanisms. Modern IAM solutions used in health care typically involve dual-factor authentication. In addition to supplying the username/passwords, users have to use an additional factor for authentication, for example, a biometric such as fingerprint, iris, or face [58]. IAM solutions also provide a feature called role-based access control, which ensures that permissions to access or modify patient data depend on the role of a person [34].

In a time-sensitive industry such as health care, the last thing clinicians want is cumbersome or repeated log-ins [38,52]. Single sign-on solutions avoid this, as the user needs to authenticate only once when accessing hospital services, rather than separately authenticating for each hospital app [34,58].

**Endpoint Security Tools**

BYOD devices need to be secured within as well as outside the hospital network. Therefore, endpoint security tools such as antivirus, antimalware, antispyware, or antiphishing tools need to be installed and regularly run on BYOD devices within hospital containers to protect and isolate hospital data within the device [52].

**Secure Clinical Communication Platforms**

Secure health care communication platforms provide an extra layer of security through strong encryption [37,59]. The United States has developed a national encryption standard called Direct for secure exchange of health care data, which provides guidelines on safe, scalable, and standards-based clinical communication and also ensures strict compliance to HIPAA [60].

**Other Emerging Technologies**

Several technologies have the potential to revolutionize the BYOD security management process. Unified endpoint management (UEM) is considered to be a good MDM alternative as it provides a single unified interface for managing all types of devices existing within the enterprise, such as PCs, laptops, smartphones, tablets, IoT devices, and wearables, which include both BYOD and company-owned devices. It also allows better methods of managing hospital apps/data, confining them to a secure workspace and separating the personal data of caregivers, which ensures the privacy of both personal health information (PHI) and personal data [61,62].

Another important technology that is gaining adoption is cloud access security broker (CASB), which is used in cloud-based MDM platforms and allows an organization to extend its security policies even outside its infrastructure and therefore manage the organizational data on the device even outside organizational parameters [63]. Health care can significantly benefit from this technology, given the mobile nature of its workforce comprising people who may work with different hospitals or health care organizations. Secure web gateway is another emerging technology that ensures that unsecured traffic, which may be initiated from BYOD devices such as malicious traffic from the web, viruses, or malware, does not enter the internal network of an organization [63]. As these technologies are relatively new, they still have not seen widespread adoption, but organizations are seriously considering their procurement [64].

**Hospital Bring-Your-Own-Device Policy Solutions**

Policy provides the required strategies, rules, and guidelines for the implementation of BYOD. The following policy components form an important part of the BYOD program.

**Bring-Your-Own-Device Strategy and Governance**

Given the lack of direction in hospitals regarding BYOD security, hospitals need to define a comprehensive hospital-wide strategy for BYOD, which should be regularly reviewed and updated. This strategy should be aligned with the hospital’s core values, future vision, and needs and requires close collaboration among all relevant stakeholders, including both clinical and nonclinical staff [65,66]. This strategy must take into consideration previous relevant procedures, data flows, and clinical workflows to understand what hospital data may be stored or transmitted from the clinician’s devices. The hospital must define who will have access to what information and where. It must also ensure that clinical productivity is not hampered [34].

**User Agreement**

Before joining the BYOD program, employees must be asked to sign a BYOD user agreement that elucidates the responsibilities of employees, defines penalties in case of noncompliance, and highlights the responsibilities of the hospital as well. An example of a BYOD user agreement is Queensland Health’s BYOD self-managed service policy document available on the web [67].

https://mhealth.jmir.org/2020/6/e18175
**Bring-Your-Own-Device Policy**

Before the rollout of the BYOD program in the hospital, policies that are in line with the BYOD strategy need to be put in place. Important elements that should be included in a BYOD policy are mentioned in Table 2 [33,34,58,65,68].

**Health care examples that exhibit these elements include Queensland Health’s BYOD policy document and the sample BYOD policy by the UK NHS [67,69].**

<table>
<thead>
<tr>
<th>Table 2. Key elements of a hospital’s bring-your-own-device policy.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item</strong></td>
</tr>
<tr>
<td>Key definitions</td>
</tr>
<tr>
<td>Service provision</td>
</tr>
<tr>
<td>Access control</td>
</tr>
<tr>
<td>Data storage</td>
</tr>
<tr>
<td>Incident reporting</td>
</tr>
<tr>
<td>Legislation and noncompliance</td>
</tr>
<tr>
<td>Education strategy</td>
</tr>
<tr>
<td>Acceptable use</td>
</tr>
</tbody>
</table>

\(^a\)BYOD: bring-your-own-device.

\(^b\)IT: information technology.

**Hospital Bring-Your-Own-Device People Solutions**

People form a critical part of the BYOD security process. The following are important measures that help improve the user security behavior of hospital employees.

**Security Culture**

All groups of hospital employees, which include the hospital’s senior management, the IT department, and BYOD users (both clinical and nonclinical staff), should be actively consulted throughout the duration of the BYOD program and made aware of their responsibilities. The leadership should make security an organizational priority so that clinicians understand the value of preserving the privacy of sensitive patient data. These steps will help in establishing a safe and secure BYOD culture, where the privacy of hospital data is taken seriously [34,65].

**Employee Awareness and Training**

Modes of training can include classroom training, computer-based training, staff meetings, monthly newsletters, posters, and regular team discussions [41]. A study conducted in 6 US health care organizations over 8 years (2011-2018) highlights the effectiveness of antiphishing campaigns in improving the security behavior of health care professionals [50].

**Skills Improvement**

Table 3 provides a summary of hospital BYOD security challenges and solutions.

Experts advocate for government investment in cybersecurity education and research and incorporation of practical training as part of academic programs. Employers also need to support their employees to complete cybersecurity certifications [70].
**Table 3. Summary of hospital bring-your-own-device security challenges and solutions.**

<table>
<thead>
<tr>
<th>Technology</th>
<th>Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>People, policy, and technology dimension and challenges</strong></td>
<td><strong>Solutions</strong></td>
</tr>
<tr>
<td><strong>Technology</strong></td>
<td></td>
</tr>
<tr>
<td>Weak authentication mechanisms</td>
<td>• Identity and access management/MDM(^a) to manage user authentication centrally</td>
</tr>
<tr>
<td></td>
<td>• Strong passwords</td>
</tr>
<tr>
<td></td>
<td>• Two-factor authentication with single sign-on</td>
</tr>
<tr>
<td></td>
<td>• Automatic log off after periods of inactivity</td>
</tr>
<tr>
<td>Malicious medical apps downloaded on BYOD(^b)</td>
<td>• Internal/regulated app stores</td>
</tr>
<tr>
<td>devices</td>
<td>• Whitelist/blacklist apps using MDM</td>
</tr>
<tr>
<td>BYOD devices connected to unsecure networks/hotspots</td>
<td>• Over-the-air network scanning</td>
</tr>
<tr>
<td></td>
<td>• Remote access through virtual private network</td>
</tr>
<tr>
<td></td>
<td>• Data protection in rest and motion (use of AES(^c)/TLS(^d))</td>
</tr>
<tr>
<td>Vulnerable devices connected on hospital network</td>
<td>• MDM to prevent vulnerable devices from connecting to hospital networks</td>
</tr>
<tr>
<td></td>
<td>• Network scanning</td>
</tr>
<tr>
<td>Mixing of personal and hospital data</td>
<td>• Containerization for logical separation of hospital and personal data</td>
</tr>
<tr>
<td></td>
<td>• Use sandboxed apps for PHI(^e) access</td>
</tr>
<tr>
<td></td>
<td>• Use secure and encrypted clinical communication platforms</td>
</tr>
<tr>
<td>Lost device containing sensitive PHI</td>
<td>• Use MDM to track/lock device remotely</td>
</tr>
<tr>
<td></td>
<td>• Use MDM with containerization to selectively wipe hospital data</td>
</tr>
<tr>
<td></td>
<td>• Limit storage of hospital data on device using virtual desktop infrastruc-</td>
</tr>
<tr>
<td></td>
<td>• Report theft incidents to hospital information technology department</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td></td>
</tr>
<tr>
<td>Lack of strategy/direction for ideal BYOD use</td>
<td>• Define hospital-wide BYOD strategy to be updated regularly</td>
</tr>
<tr>
<td></td>
<td>• Dedicated BYOD policy for complete guidance on authentication, access con-</td>
</tr>
<tr>
<td></td>
<td>• chain of responsibility, data ownership, devices allowed, acceptable use,</td>
</tr>
<tr>
<td></td>
<td>• training, legislation, and noncompliance</td>
</tr>
<tr>
<td></td>
<td>• Mandating signing of user agreement for BYOD users</td>
</tr>
<tr>
<td>Maintaining compliance with health care data protec-</td>
<td>• Notify relevant government departments about breaches as per law</td>
</tr>
<tr>
<td>tion laws</td>
<td>• Perform regular audits and legal risk assessments</td>
</tr>
<tr>
<td></td>
<td>• Define applicable privacy regulations and penalties for noncompliance</td>
</tr>
<tr>
<td></td>
<td>• Train BYOD users about incident reporting to notify breaches/thefts</td>
</tr>
<tr>
<td>Access privilege abuse</td>
<td>• Use principle of least privileges and role-based access control in defining staff access to PHI</td>
</tr>
<tr>
<td><strong>People</strong></td>
<td></td>
</tr>
<tr>
<td>Inappropriate behavior by BYOD users</td>
<td>• Penalize staff found guilty of breaches</td>
</tr>
<tr>
<td></td>
<td>• Encourage safe and secure use by establishing a security culture</td>
</tr>
<tr>
<td></td>
<td>• Monitor user behavior regularly</td>
</tr>
<tr>
<td>Lack of awareness among hospital BYOD users</td>
<td>• Educate BYOD users periodically</td>
</tr>
<tr>
<td></td>
<td>• Check awareness levels regularly, for example, through phishing campaigns</td>
</tr>
<tr>
<td>Poor user experience</td>
<td>• Consult all relevant stakeholders throughout the BYOD program</td>
</tr>
<tr>
<td></td>
<td>• Carefully consider clinical workflow and ease of use</td>
</tr>
<tr>
<td>Cybersecurity budget and skills shortage</td>
<td>• Government investment in technology, education, and research</td>
</tr>
<tr>
<td></td>
<td>• Hiring experts</td>
</tr>
<tr>
<td></td>
<td>• Sponsoring and supporting employees for skills improvement</td>
</tr>
</tbody>
</table>

\(^{a}\)MDM: mobile device management.  
\(^{b}\)BYOD: bring-your-own-device.  
\(^{c}\)AES: Advanced Encryption Standard.  
\(^{d}\)TLS: Transport Layer Security.  
\(^{e}\)PHI: personal health information.
Discussion

Principal Findings

A wide range of technological solutions, policy control measures, and social practices are available that can be used together to curb BYOD security risks in hospitals. The findings suggest that the key challenge lies in ensuring a proper balance between usability and security. Therefore, BYOD security management should not only involve the use of resilient security mechanisms but also ensure that the mobility and productivity of hospital employees, especially clinicians, are not hampered. Hospital-owned enterprise productivity apps such as secure messaging and photography or the use of single sign-on for accessing hospital apps can help to address such concerns.

From a technological perspective, the BYOD landscape is changing. Gartner analysts predict that the BYOD model will change to a bring-your-own-environment model as users bring not only devices and apps but also services, personal networks, and even collaborative workspaces beyond email or messaging [71]. It is expected that large health care organizations will have to deal with as many as 80,000 connected IoT devices [72]. Modern BYOD security technologies such as UEM or containerization allow BYOD security management to become device independent. The highlight of these technologies is that they allow logical separation of hospital and personal data on employees’ devices. As such, BYOD security management is moving from a traditional device-centric approach to a data-centric or app-centric approach, to gain control of hospital data residing on employees’ devices and relinquish the management of the whole device itself.

On the basis of our findings from the policy perspective, this study highlights how a lack of clarity with regard to optimal BYOD usage prevalent in hospitals can be addressed through a dedicated BYOD policy as it will provide complete direction and guidelines for safe, secure, and productive BYOD usage in the hospital. Important components that must be covered in the policy include governance of the BYOD program, choosing the right technologies to support the program, providing guidelines for appropriate use, training and education strategies for employees, and compliance with health care data privacy laws. In addition, the policy needs to be contextual and inclusive of various stakeholders. The policy must also be in alignment with the organization’s goals and the limits of the hospital’s capabilities. Finally, with rapid changes in BYOD technology, modernization of policy is also required. This means that the introduction or reformation of security standards or protocols needs to be contemplated by considering both security and the clinical workflow [73].

From people’s perspective, effective implementation of technology and policy will require strong consultations with all relevant stakeholders from the management side, which includes the hospital’s senior management; the IT department; and the user side, including BYOD users such as clinicians, administrative staff, and other staff. Educating and training employees about BYOD threats and security measures should be directed to improve their commitment to protect hospital data, especially PHI. The changing BYOD landscape also demands expansion and refreshment of skill set for the management, IT department, and BYOD users. Hospitals may need to recruit external staff capable of successfully implementing complex technologies such as UEM and CASB, which may also require training of hospital IT personnel [62].

Limitations

As far as the limitations of this study are concerned, the complex nature of gray literature searches means that the number of items available for review is significantly higher than that of peer-reviewed literature searches; therefore, limiting the items analyzed was unavoidable [24]. Nevertheless, some important sources may have been neglected. Furthermore, although some gray literature sources gave examples from real-life studies, others were based on the opinions of credible experts who wrote blogs for well-known media publications. Although these sources are useful, they do not meet the criteria of the highest quality sources.

Comparison With Prior Work

The use of the PPT model [27] to answer the research questions has aided in providing a holistic perspective of BYOD security management, which may have been lacking in previous studies, as highlighted earlier. This study, therefore, complements and augments the authors’ previous findings, where the same model was used for analyzing peer-reviewed literature [17]. The practical, real-life evidence extracted through the gray literature review not only corroborates the previous outcomes but also provides additional insights.

Conclusions

As modern BYOD security threats grow in size and complexity, this study elucidates how health care organizations can use technological solutions, policy control mechanisms, and people management measures in close alignment to curb such risks effectively and holistically. This has become very important as cybersecurity is seen as one of the biggest challenges in the health care industry [74], with BYOD being one of the major threats to cybersecurity itself [39,75].

Acknowledgments

This work was funded by the School of Computing and Information Systems and the Centre for Digital Transformation of Health at the University of Melbourne. The authors would also like to acknowledge the University of Melbourne’s support through its PhD research scholarship.

https://mhealth.jmir.org/2020/6/e18175

JMIR Mhealth Uhealth 2020 | vol. 8 | iss. 6 | e18175 | p.471
(page number not for citation purposes)
Conflicts of Interest
None declared.

Multimedia Appendix 1
Gray literature summary and characteristics.
[PDF File (Adobe PDF File), 179 KB - mhealth_v8i6e18175_app1.pdf ]

References


Abbreviations

BYOD: bring-your-own-device  
CASB: cloud access security broker  
HIMSS: Healthcare Information and Management Systems Society  
HIPAA: Health Insurance Portability and Accountability Act  
IAM: identity and access management  
IoT: internet of things  
IT: information technology  
MDM: mobile device management  
NHS: National Health Services  
PHI: personal health information  
PPT: People Policy Technology  
UEM: unified endpoint management  
VPN: virtual private network

©Tafheem Ahmad Wani, Antonette Mendoza, Kathleen Gray. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 18.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Abstract

Background: With the unprecedented growth of mobile technology, a plethora of dialysis diet apps have been developed to promote patient dietary self-management. Nevertheless, the utility of such apps remains questionable.

Objective: This study aimed to evaluate the content, features, and quality of commercial dialysis diet apps for adult dialysis patients.

Methods: This study consisted of a quantitative content analysis of commercial dialysis diet apps downloaded from Google Play and the Apple App Store available in the Asian marketplace, searched for using the following keywords in English: dialysis diet and diet for kidney disease. Free and paid apps available in English that provide nutrition information for adult dialysis patients were included. Apps that were not relevant to the dialysis diet, not meant for patient self-management, or redundant were excluded. Apps were evaluated for language medium (subscore=1), credibility (subscore=1), food database (subscore=1), valuable features (subscore=12), health-behavior theory constructs (subscore=60), and technical quality (subscore=25). The relationships among the variables of interest were determined by Pearson correlation. Stepwise multiple linear regression analysis was performed to identify the features that contribute to greater technical quality of the commercial dialysis diet apps. Statistical significance was defined as P<.05.

Results: A total of 22 out of 253 apps (8.7%) were eligible for evaluation. Based on a 100-point scale, the mean overall score of the apps was 31.30 (SD 14.28). Only 5% (1/22) of the apps offered relevant language options, and 46% (10/22) contained food databases. In addition, 54% (12/22) of the apps were not credible. The mean score for valuable features was 3.45 (SD 1.63) out of 12, in which general education (16/22, 73%), free download (15/22, 68%), and usability (13/22, 59%) were the three most popular features. However, the apps scored a mean of 13.41 (SD 11.56) out of 60 for health-behavior theory constructs. The overall app technical quality was considered poor, with a mean score of 2.70 (SD 0.41) out of 5. The scores of valuable features (r=.65, P=.001) and health-behavior theory constructs (r=.55, P=.009) were positively correlated with the overall technical quality of the commercial dialysis diet apps. Features such as free download (β=.43, P=.03) and usability (β=.41, P=.03) could significantly determine the functional quality of the apps. Health-behavior theory constructs such as self-monitoring could significantly predict both the subjective quality (β=.55, P=.008) and the engagement quality (β=.66, P=.001) of the apps, whereas the information quality domain could be determined by plan or orders (β=.48, P=.007) and knowledge (β=.45, P=.01).
Conclusions: Although most of the available commercial dialysis diet apps are free and easy to use, they are subject to theory deficiency, limited language options, and a lack of food databases, credibility, tailored education, and overall technical quality.

KEYWORDS
renal apps; nutrition; dialysis; self-management; mHealth

Introduction

With the unprecedented growth of mobile technology, the use of mobile phones is ubiquitous around the globe. Such usage has been proliferating over the years, with a world penetration rate of 67% in 2019 [1]. In this modern era, mobile phones are not only communication tools but also indispensable devices that enable users to perform a variety of activities, such as those related to entertainment, social media, fitness, and health care.

Recently, there has been growing interest in mobile health (mHealth) app development. According to a survey [2], approximately 58,000 mHealth app publishers existed in 2016. Among various categories, nutrition-related apps were found to be the most downloaded mHealth apps [3]. The positive effects of nutrition-related mHealth apps as self-monitoring tools in managing chronic diseases, particularly concerning weight management, have been supported by a recent systematic review and meta-analysis [4]. Moreover, the same study has also addressed a gap pertaining to the effects of dietary mHealth apps on chronic kidney diseases (CKDs).

Diet modification is one of the most crucial components of comprehensive dialysis treatment [5]. Poor dietary adherence will result in life-threatening complications [6-8]. However, diet modification for dialysis patients is challenging due to the complexity of the renal diet. Such diet modification requires a substantial amount of patient self-management skills to integrate and implement the complex dietary recommendations over the course of one’s lifetime [9].

A plethora of dialysis diet apps are now available in mobile app stores. Their roles as dietary self-management tools (ie, diet trackers, food diaries, calorie counting functions, and nutrition recommendations) for dialysis patients as adjuncts to dietetics counseling have been increasingly advocated [10,11]. However, there is no strong conclusive evidence to support the clinical efficacy of these apps [12]. Poor app engagement and usability issues are believed to be the reasons for the limited utility of these apps [11]. In addition, due to the absence of consensus standards and development guidelines for mHealth apps [13], the quality of commercial dialysis diet apps is questionable.

An earlier study found that approximately half of the commercial diet apps for kidney diseases were not credible and only had fair technical quality [14]. Nevertheless, other important aspects of commercial dialysis diet apps that may be linked to greater user acceptance, engagement, and effectiveness have yet to be explored, such as health behavioral theory [15], a set of valuable features [16], and other aspects, such as language medium [17] and food databases [18].

Therefore, this study aims to examine the content, features, and quality of commercial dialysis diet apps to inform health care professionals and patients about the current state of dialysis diet apps and to address the aspects of apps that must be improved.

Methods

Study Design

In this study, we performed a quantitative content analysis of commercial dialysis diet apps from the two most popular mobile platforms in the Asian marketplace: Google Android (ie, downloaded from the Google Play store) and Apple iOS (ie, downloaded from the Apple App Store). The content, features, and quality of the eligible apps were evaluated and quantified using a predefined scoring system. Considering the nature of the study design (ie, desk-based study), ethical approval was exempted from this study.

Sampling Method

Dialysis diet apps were sampled from Google Play and the Apple App Store for the Asian marketplace using plausible keywords that dialysis patients would use to search for renal diet apps (ie, dialysis diet and diet for kidney disease). The search was conducted by two research staff members from September 26, 2018, to October 31, 2018. The apps identified in the initial search were screened for eligibility using predefined selection criteria. The inclusion criteria for apps in this study were as follows: apps were free or paid, were available in English, and provided nutrition information, including fluid control for adult dialysis patients. In contrast, apps were excluded if they were not relevant to the dialysis diet, not meant for patient self-management, or redundant. Screening was conducted independently by two study staff members to minimize researcher bias. Discrepancies in the screening results were compared and discussed between the staff members before a final list of dialysis diet apps was constructed for further evaluation.

App Evaluation

Eligible dialysis diet apps were evaluated concerning six main aspects. These included the following: (1) the presence of valuable features, (2) the extent of health-behavior theory incorporation, (3) credibility, (4) technical quality, (5) the option of a language medium, and (6) the presence of a food database. The evaluation aspects included in this study were chosen based on the extant literature pertaining to the features and characteristics of mHealth apps that might be associated with greater user acceptance, engagement, and effectiveness [14-18].

https://mhealth.jmir.org/2020/6/e13808
Scoring System

Overview
A scoring system encompassing all evaluation aspects was developed to quantify the evaluation outcomes. The scoring system consisted of a rubric and scale adopted from the literature [15,16,19]. Apps were scored if they fulfilled the predetermined criteria according to the scoring distribution: language medium (subscore=1), credibility (subscore=1), food databases (subscore=1), valuable features (subscore=12), health-behavior theory constructs (subscore=60), and technical quality (subscore=25). The scoring distribution was constructed as the number of components within each criterion and/or their respective scoring scale. For instance, the subscore assigned to the aspect of valuable features was 12 because it consists of 12 evaluation features [16], as shown in Multimedia Appendix 1. Health-behavior theory [15] and technical quality [19] were scored according to their respective scales as described above. Since credibility, language, and food databases are stand-alone features, they only contributed a score of 1 each to the overall score. The overall score of each app ranged from 0 to 100, with a higher score indicating higher quality in terms of content and features that were thought to be linked to a greater acceptance, engagement, and effectiveness of the mHealth app.

Valuable Features
Eligible apps were assessed for the presence of valuable features adopted from the previous content analysis of mHealth apps [16]. There were 12 valuable features found to be associated with user engagement and positive user rating [16]: (1) export of data, (2) gamification, (3) general education, (4) plan or orders, (5) reminders, (6) community forum, (7) social media, (8) addressing of symptoms, (9) tailored education, (10) tracker, (11) cost (ie, free download), and (12) usability. The description of each feature is presented in Multimedia Appendix 1. Using a binary system, apps were given a score of 1 to indicate the presence of a specific feature. Otherwise, a score of 0 was given. Eventually, the score of each feature was summed, with the overall subscore ranging from 0 to 12.

Health-Behavior Theory
A rubric utilized by previous studies [15,20] was adopted to examine the extent of incorporating health-behavior theory constructs into commercial dialysis diet apps. Eligible apps were assessed for the presence of 12 theoretical constructs (ie, knowledge, perceived benefits, perceived barriers, perceived risks, self-efficacy, social norm, self-monitoring, goal setting, stimulus control, self-reward, social support, and vicarious learning), grounded from the four most commonly used health-behavior theories in health apps, including (1) the health belief model, (2) the transtheoretical model, (3) the theory of planned behavior, and (4) social cognitive theory. Then, each of the constructs was coded based on six levels of user interaction. The construct was rated as 0 if user interaction was absent, 1 if it involved general information, 2 for assessment, 3 for feedback, 4 for general assistance, and 5 for individually tailored assistance (see Multimedia Appendix 2). The subscores for this section ranged from 0 to 60 (ie, 12 constructs x six levels of user interaction), with a higher score indicating a greater extent of the incorporation of health-behavior theory.

Credibility
The content of the eligible apps was examined by comparing it with the dietary recommendations for the adult CKD population [21], which were derived from numerous nutrition guidelines [22-26]. Apps with inconsistent information or without reliable references were labeled as not credible.

Conversely, for apps that function solely as diet trackers, their credibility was assessed primarily through the accuracy of their food databases. A total of 20 food items (ie, four food items per major food group: cereals, protein, vegetables, fruits, and fat), which represent an average number of food items consumed by a person per day [27], were randomly selected for comparison using cross-classification analysis, which is a widely used method to assess agreement in validation studies involving food nutrients [28,29]. Apps that showed more than 7% gross misclassification [30] were considered not credible. In the absence of a gold standard, the nutrient content of food in the databases of the dialysis diet apps was compared with that of the computer software Nutritionist Pro, version 2.2.16 (First DataBank Inc) [31]. Nutritionist Pro has been widely utilized to assess dietary intake in many published research studies [32-34], including for CKD [35-37]. In addition, it has also been used to analyze dietary data in numerous validation studies on dietary assessment tools, such as food frequency questionnaires [38,39]. Nutritionist Pro contains food databases of various regions, such as the United States Department of Agriculture, the Canadian Nutrient File, and UK, European, and Malaysian food databases. As the only food database for Asian countries in Nutritionist Pro, the Malaysian food database was used in this study to determine the credibility of the apps.

In this study, apps that provided inaccurate or partially accurate information or were not from reliable sources were given a score of 0. In contrast, a score of 1 was given to credible apps.

Technical Quality
The technical quality of the apps was evaluated using the Mobile Application Rating Scale (MARS) [19], which is a validated tool specifically designed to evaluate the technical quality of mHealth apps based on five quality domains: (1) engagement (ie, entertainment, interest, customization, interactivity, and fit to target group), (2) functionality (ie, performance, ease of use, navigation, and gestural design), (3) aesthetics (ie, layout, graphic design, and visual appeal), (4) information (ie, quality and quantity of visual information, credibility, goal, and accuracy of app description), and (5) subjective quality (ie, recommendation, willingness to pay, willingness to use in future, and overall satisfaction). In this study, the MARS has been adopted without any modification to its items and domains to evaluate the technical quality of the apps. However, to fulfill the needs for this study, the scoring method has been adapted by summing the subscores of all domains. Each question was scored on a scale of 1 (inadequate) to 5 (excellent), with the total score ranging from 5 (minimum) to 25 (maximum). The validity and reliability of the MARS with the revised scoring method has been determined using the evaluated apps. Good
Good interrater reliability was determined, with kappa and ICC values equal to .798 and .762, respectively. Discrepancies in the rating were discussed until a consensus was reached. Then, each eligible app was downloaded and coded independently by the raters for approximately 1 week. The final score of each app was obtained by averaging the scores from the raters.

**Statistical Analysis**

Descriptive statistics were used to summarize the findings of this study. All data were analyzed using SPSS, version 25.0 (IBM Corp). Categorical variables were expressed as frequencies and percentages, while continuous data were expressed as the mean (SD). The relationships among variables of interest were determined by Pearson correlation. Stepwise multiple linear regression analysis was performed to identify the features that predict the technical quality of apps. Statistical significance was defined as \( P < .05 \).

**Results**

**Search Results**

A total of 253 apps were identified in the initial search. However, only 8.7% (22/253) of the apps were eligible for evaluation. Of these apps, 73% (16/22) were Android-based apps, while the remaining 27% (6/22) were Apple iOS apps. A large proportion of the apps (231/253, 91.3%) were excluded, as they were not relevant to dialysis diets, such as non-dialysis-specific diet apps (ie, fitness, diabetes, uric acid, kidney stones, etc) and calculators (ie, glomerular filtration rate calculator), which did not provide any dialysis-related diet information and were not meant for patient self-management (ie, journals, medical pocketbooks, etc). Other reasons for exclusion were redundant apps, apps available only in other languages (ie, Urdu, German, and Spanish), and apps that could not be downloaded or used. The sampling details are presented in Figure 1.
Figure 1. Flow diagram of the selection process and filtering results for the content analysis of existing dialysis-specific diet apps.

<table>
<thead>
<tr>
<th>Evaluation Outcomes</th>
</tr>
</thead>
</table>

Table 1 depicts the summary scores of the app evaluation. Based on the 100-point scale, the mean overall app score was 31.30 (SD 14.28), ranging from 10.28 (lowest) to 53.82 (highest). However, most of the apps (19/22, 86%) scored less than 50 points. The mean score acquired by Android-based apps was 35.13 (SD 13.18), while Apple iOS-based apps obtained an average score of 21.09 (SD 12.79). The scores of the evaluated apps are presented in Multimedia Appendix 3.

Only 5% (1/22) of the apps offered language options relevant to the Asian marketplace, while approximately 46% (10/22) contained food databases (see Table 1). In addition, 54% (12/22) of the evaluated apps were not credible, and Android-based apps (9/16, 56%) were more credible than Apple iOS-based apps (1/6, 17%). Commercial dialysis diet apps scored 3.45 (SD 1.63) out of 12 for valuable features; Android-based dialysis diet apps (3.88, SD 1.59) contained more valuable features than did Apple iOS-based dialysis diet apps (2.33, SD 1.21), as shown in Table 1.

The mean overall MARS score of commercial dialysis diet apps was 13.48 (SD 2.05) out of a total score of 25 (see Table 1). Android dialysis diet apps (14.04, SD 1.87) have better technical quality than Apple iOS apps (12.01, SD 1.92). The mean score of commercial dialysis diet apps across the five MARS quality domains was 2.70 (SD 0.41) out of 5, with the highest score being for functionality (3.79, SD 0.45), followed by those for aesthetics (2.95, SD 0.44), engagement (2.40, SD 0.66), information (2.27, SD 0.59), and subject quality (2.08, SD 0.48).

The presence of valuable features in commercial dialysis diet apps is depicted in Multimedia Appendix 4. The app that contained the highest number of valuable features (ie, seven) was the Android-based app Renal Care Compass. In contrast, the Apple iOS app Healthy Kidneys Grocery List had the least valuable features (ie, one). The three most popular valuable features found in commercial dialysis diet apps were general education (16/22, 73%), followed by free download (15/22, 68%) and usability (13/22, 59%). Moreover, features such as gamification (1/22, 5%), tailored education (1/22, 5%), social media (0/22, 0%), and community forums (0/22, 0%) were the least incorporated features in commercial dialysis diet apps.

Out of a total score of 60, the mean score of commercial dialysis diet apps for health-behavior theory constructs was 13.41 (SD 11.56) (see Table 1). In general, Android-based dialysis diet apps (16.16, SD 11.11) applied theoretical constructs to a greater extent than did Apple iOS-based dialysis diet apps (6.08, SD 10.10). The health-behavior theory constructs integrated into commercial dialysis diet apps are presented in Multimedia Appendix 5. The Android-based dialysis diet app Renal Disease Kidney Diet Tips Symptoms & Foods was the most theory-based dialysis diet app. Surprisingly, 4 out of 6 (67%) Apple iOS-based dialysis diet apps did not integrate any theoretical construct evaluated in this study. Knowledge (17/22, 77%), goal setting (15/22, 68%), and self-efficacy (13/22, 59%) were the most widely used theoretical constructs in commercial dialysis
diet apps. In contrast, perceived barrier (2/22, 9%), self-reward (2/22, 9%), social support (1/22, 5%), and vicarious learning (1/22, 5%) were the least incorporated theoretical constructs. Collectively, the findings of the evaluation of commercial dialysis diet apps by evaluating different aspects are summarized in percentages and illustrated in Figure 2.

Table 2 presents the relationships among the valuable features, health-behavior theory, and technical quality of commercial dialysis diet apps. Except for aesthetics and functionality quality domains, valuable features were significantly correlated with overall technical quality ($r=.65, P=.001$), the engagement quality domain ($r=.45, P=.04$), and the information quality domain ($r=.53, P=.01$), but not with the aesthetics, functionality, and subjective quality domains ($P>.05$).

Stepwise regression analysis indicated that only cost ($\beta=.49$, $P=.005$) and self-monitoring ($\beta=.46, P=.009$) could significantly predict the overall quality of commercial dialysis diet apps (see Table 3). In addition, self-monitoring was also a predictor of the engagement ($\beta=.66, P=.001$) and subjective quality ($\beta=.55, P=.008$) domains. The functionality quality of the app could be determined by cost ($\beta=.43, P=.03$) and usability ($\beta=.41, P=.03$), while the information quality domain could be determined by plan or orders ($\beta=.48, P=.007$) and knowledge ($\beta=.45, P=.01$).

Table 1. Summary scores of the evaluated dialysis diet apps.

<table>
<thead>
<tr>
<th>Evaluating aspects</th>
<th>Overall (N=22)</th>
<th>Android-based apps (n=16)</th>
<th>Apple iOS-based apps (n=6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language (score out of 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score, mean (SD)</td>
<td>0.05 (0.21)</td>
<td>N/A^a</td>
<td>N/A</td>
</tr>
<tr>
<td>English only, n (%)</td>
<td>21 (95)</td>
<td>15 (94)</td>
<td>6 (100)</td>
</tr>
<tr>
<td>Multilanguage, n (%)</td>
<td>1 (5)</td>
<td>1 (6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Food database (score out of 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score, mean (SD)</td>
<td>0.45 (0.51)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Present, n (%)</td>
<td>10 (46)</td>
<td>7 (44)</td>
<td>3 (50)</td>
</tr>
<tr>
<td>Absent, n (%)</td>
<td>12 (54)</td>
<td>9 (56)</td>
<td>3 (50)</td>
</tr>
<tr>
<td><strong>Credibility (score out of 1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score, mean (SD)</td>
<td>0.45 (0.51)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes, n (%)</td>
<td>10 (46)</td>
<td>9 (56)</td>
<td>1 (17)</td>
</tr>
<tr>
<td>No, n (%)</td>
<td>12 (54)</td>
<td>7 (44)</td>
<td>5 (83)</td>
</tr>
<tr>
<td>Valuable features (score out of 12), mean (SD)</td>
<td>3.45 (1.63)</td>
<td>3.88 (1.59)</td>
<td>2.33 (1.21)</td>
</tr>
<tr>
<td>Theory constructs (score out of 60), mean (SD)</td>
<td>13.41 (11.56)</td>
<td>16.16 (11.11)</td>
<td>6.08 (10.10)</td>
</tr>
<tr>
<td><strong>Technical quality (score out of 25), mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score</td>
<td>13.48 (2.05)</td>
<td>14.04 (1.87)</td>
<td>12.01 (1.92)</td>
</tr>
<tr>
<td>Engagement</td>
<td>2.40 (0.66)</td>
<td>2.60 (0.60)</td>
<td>1.88 (0.55)</td>
</tr>
<tr>
<td>Functionality</td>
<td>3.79 (0.45)</td>
<td>3.81 (0.50)</td>
<td>3.73 (0.33)</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>2.95 (0.44)</td>
<td>3.03 (0.43)</td>
<td>2.72 (0.39)</td>
</tr>
<tr>
<td>Information</td>
<td>2.27 (0.59)</td>
<td>2.43 (0.52)</td>
<td>1.83 (0.60)</td>
</tr>
<tr>
<td>Subjective quality</td>
<td>2.08 (0.48)</td>
<td>2.17 (0.50)</td>
<td>1.84 (0.32)</td>
</tr>
<tr>
<td>Mobile Application Rating Scale score</td>
<td>2.70 (0.41)</td>
<td>2.81 (0.37)</td>
<td>2.40 (0.38)</td>
</tr>
<tr>
<td>Total score (score out of 100), mean (SD)</td>
<td>31.30 (14.28)</td>
<td>35.13 (13.18)</td>
<td>21.09 (12.79)</td>
</tr>
</tbody>
</table>

^aN/A: not applicable.
Figure 2. Radar chart for the evaluation results of existing dialysis-specific diet apps.

Table 2. Correlations\(^a\) between mean scores of technical quality with mean scores of valuable features and health-behavior theory for evaluated dialysis diet apps (N=22).

<table>
<thead>
<tr>
<th>Technical quality</th>
<th>Valuable features</th>
<th>Health-behavior theory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(r)</td>
<td>(P) value</td>
</tr>
<tr>
<td>Overall</td>
<td>.65</td>
<td>.001</td>
</tr>
<tr>
<td>Engagement</td>
<td>.60</td>
<td>.003</td>
</tr>
<tr>
<td>Functionality</td>
<td>.27</td>
<td>.22</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>.36</td>
<td>.10</td>
</tr>
<tr>
<td>Information</td>
<td>.61</td>
<td>.002</td>
</tr>
<tr>
<td>Subjective quality</td>
<td>.61</td>
<td>.003</td>
</tr>
</tbody>
</table>

\(^a\)Analyzed using Pearson correlation.
Table 3. Predictors of technical quality in evaluated dialysis diet apps (N=22)\(^a\).

<table>
<thead>
<tr>
<th>Technical quality and predictors</th>
<th>(R^2)</th>
<th>B</th>
<th>SE</th>
<th>(\beta)</th>
<th>(t) test</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>.56</td>
<td>1.85</td>
<td>.64</td>
<td>.46</td>
<td>2.90</td>
<td>.009</td>
</tr>
<tr>
<td>Cost</td>
<td>N/A(^b)</td>
<td>2.14</td>
<td>.68</td>
<td>.49</td>
<td>3.14</td>
<td>.005</td>
</tr>
<tr>
<td><strong>Engagement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>.43</td>
<td>.85</td>
<td>.22</td>
<td>.66</td>
<td>3.92</td>
<td>.001</td>
</tr>
<tr>
<td><strong>Functionality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost</td>
<td>.44</td>
<td>.41</td>
<td>.17</td>
<td>.43</td>
<td>2.42</td>
<td>.03</td>
</tr>
<tr>
<td>Usability</td>
<td>N/A</td>
<td>.37</td>
<td>.16</td>
<td>.41</td>
<td>2.33</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan or orders</td>
<td>.54</td>
<td>.72</td>
<td>.24</td>
<td>.48</td>
<td>3.00</td>
<td>.007</td>
</tr>
<tr>
<td>Knowledge</td>
<td>N/A</td>
<td>.62</td>
<td>.22</td>
<td>.45</td>
<td>2.90</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Subjective quality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-monitoring</td>
<td>.31</td>
<td>.52</td>
<td>.18</td>
<td>.55</td>
<td>2.97</td>
<td>.008</td>
</tr>
</tbody>
</table>

\(^a\)Analyzed using stepwise multiple linear regression.

\(^b\)N/A: not applicable.

**Discussion**

**Principal Findings**

Based on the findings of this study, only a limited number of commercial renal diet apps (22/253, 8.7%) are available for dialysis patients. Moreover, these apps were found to be lacking in language options relevant to Asian marketplaces (1/22, 5%) and food databases (10/22, 46%). They also have poor technical quality (mean 13.48, SD 2.05, out of 25) associated with limited valuable features (mean 3.50, SD 1.68, out of 12) and health-behavior theory incorporation (mean 13.41, SD 11.56, out of 60).

Renal patients have shown a growing interest in using mHealth apps [41]. Unfortunately, despite having numerous renal diet apps in mobile app stores, only a limited number of apps are likely to fulfill the needs of dialysis patients, regardless of their quality. In addition, renal apps that are meant for patient self-management are still limited [42]. Choices of dialysis diet apps are further limited by the absence of language options, as not all patients are literate in the English language, especially in most Asian countries [40]. In this study, only one app (ie, *Chronic Kidney Disease*) offered language options relevant to Asian marketplaces, including Mandarin, Thai, Indonesian, Korean, and Japanese languages. Moreover, although language options (ie, Catalan and Spanish) were found in the app *Pukono*, these languages are not relevant to the Asian marketplace. The implications of the possible language barrier in mHealth apps have been discussed in a previous study [43]. The study showed that most mobile phone users prefer to use apps in their primary language [44], which can lead to greater user engagement and prevent the misinterpretation of health information.

This study found that the overall technical quality of commercial dialysis diet apps assessed by the MARS was poor. Functionality was found to be the top-rated quality domain of the MARS, associated with the usability of apps. Generally, commercial dialysis diet apps work well with minimal technical errors. However, serious usability problems, such as lagging, were detected in certain apps (ie, *Chronic Kidney Disease*), which may cause frustration among their users [45]. In addition to app performance, functionality also refers to user experience. Both dietitians and patients prefer an app that is simple and intuitive to use [46,47]. Although the majority of the commercial dialysis diet apps (13/22, 59%) were rated as easy to use, features that allow for easy control interactions were absent in certain apps. For instance, users must fill in their basic information (ie, age, gender, and height) whenever logging in to the apps. In addition, important gestural designs, such as pinch for zooming, were also absent in certain apps (ie, *Renal TRKRR*). Considering the possible vision problems in dialysis patients secondary to aging or concomitant disease (ie, diabetic retinopathy), the content of the apps might be too small to be seen.

A low retention rate remains a critical issue with mHealth apps [48]. Commercial dialysis diet apps are lacking in interactive features (ie, feedback), making the apps less engaging to users. Although gamification is a trending feature for promoting user engagement [49], it is one of the least exploited features in commercial dialysis diet apps. Based on our findings, the engagement quality of apps assessed by the MARS can be improved by incorporating self-monitoring features, as they promote user interaction (ie, diet tracking and feedback) [50]. Although 46% (10/22) of commercial dialysis diet apps offer self-monitoring features, the food items available in the database are mainly Western foods. This might limit their usefulness, especially for users from non-Western countries. A local food database is necessary to provide accurate dietary self-monitoring.
Dietary assessment using a foreign food database may cause a significant error during dietary assessment [14]. As expected, the evaluated apps scored poorly on the information quality domain of the MARS due to the lack of accuracy in the health information provided. Credibility is a prerequisite for useful mHealth apps, and it is always the greatest concern of health care professionals [46]. More than half of the commercial dialysis diet apps evaluated in this study were not credible, which may create uncertainties among health care professionals in recommending mHealth apps to their patients. From the patient's perspective, unreliable health information in the apps can bring serious detrimental effects, exacerbating the health of this vulnerable population [10]. The lack of input from health care professionals was regarded as the main reason for misinformation in commercial mHealth apps [52]. Thus, the involvement of health care professionals (ie, renal dietitians and nephrologists) in mHealth app development is advocated to ensure the credibility of the health information given [14]. Furthermore, dialysis diet apps tend to have better information quality if they provide information and guidance about renal diet (ie, dietary plan and knowledge) compared to those that function solely as a tracker (ie, Renal TRKRR). Moreover, none of the evaluated diet apps in this study were previously tested in a clinical trial. Although a previous study found potential clinical benefits of renal diet apps in the dialysis population [12] (ie, BalanceLog and Dietary Intake Monitoring Application [DIMA]), they were not available in the mobile app stores of Google and Apple during the evaluation period.

The aesthetic quality domains of the evaluated apps in this study were below average. Of these, the color, design, and layout of commercial dialysis diet apps need to be improved. Although color is not a primary concern when designing an app, it exerts a profound effect on user experience and overall satisfaction [53]. Poor color combination, especially the background color, may affect the readability of the text [54]. Moreover, the evaluated apps scored the lowest for the domain of subjective quality in the MARS. Approximately 46% (10/22) of the apps were rated below average, and minimal usage (ie, less than two times per year) was expected for most apps (18/22, 82%). The raters involved in this study represent the perspective of dietitians on commercial dialysis diet apps. Out of 22 apps, only 7 (36%) were likely to be recommended to dialysis patients. In addition, we found that dietitians (ie, raters) prefer apps with self-monitoring features that allow dialysis patients to monitor their nutrient intake.

The cost of apps is also an important criterion to be considered. Although paid apps are always deemed to be superior to free apps [16], it is not the sole indicator of better quality for commercial dialysis diet apps. In this study, we found that free dialysis diet apps outperformed paid apps in almost every evaluated aspect. The possible reason is that rather than offering additional features and functions, paid apps are generally meant for better user experience (ie, ad free) [16]. This is supported by the features offered by in-app purchases. Only a small subset (2/22, 9%) of the evaluated apps offer in-app purchases in this study. However, they are mainly used to avoid advertisements and do not contribute to any additional features evaluated. This may also explain the finding that Android apps outperform iOS apps, as most of the iOS dialysis diet apps (5/6, 83%) are paid apps compared to Android dialysis diet apps (2/16, 13%).

Health-behavior theory plays a crucial role in mHealth apps [55,56]. Similar to previous studies [15,57], the constructs of social cognitive theory (ie, knowledge, goal setting, and self-efficacy) were the most common theoretical constructs found in commercial dialysis diet apps. They have been used in designing mHealth app interventions for chronic diseases, including hemodialysis populations [58]. Consistent with previous findings [15], the extent of theory application in commercial dialysis diet apps was restricted to general information and general assistance, which are considered insufficient to bring about significant and long-term behavior change [59]. Instead, individualized dietary feedback based on assessments is more likely to promote sustainable behavior changes [60]. In this study, only one app (ie, Renal Disease Kidney Diet Tips Symptoms & Foods) was found to provide individualized assistance through social support (ie, online consultation).

Strengths and Limitations
Since Google Android and Apple iOS are the most popular mobile platforms worldwide, the findings of this study can serve as a reference for global health care professionals and the dialysis population. However, this study has several limitations. The commercial dialysis diet apps included in this study were searched for over a short period (ie, September 26 to October 31, 2018) using English keywords only. In addition, they were also confined to the apps available for Google Android and Apple iOS for the Asian marketplace only. Thus, the findings cannot be inferred for apps launched after the study period and those available in other platforms and languages.

Comparison With Prior Work
Prior studies had been conducted to evaluate commercial renal apps designed for both renal patients and health care professionals. These include studies pertaining to diet apps for general kidney diseases [14] and health apps specific to CKD management [42,61]. In comparison, this study was focused on diet apps designed specifically for dietary self-management in the dialysis population. In addition, the use of different mobile platforms as well as the keywords used to identify apps resulted in a different number of renal apps being evaluated. Moreover, this study evaluated apps on different aspects compared to previous studies, which focused more on the functionality and content of the apps rather than their language medium, food database, presence of valuable user features, and incorporation of health-behavior theory.

Despite different app pools, our findings were consistent with a previous study in which 45.5% of renal diet apps were found to be not credible [14]. Similarly, functionality was found to be the top technical quality domain assessed by the MARS. In contrast, the overall mean technical quality of dialysis-specific diet apps in this study was slightly lower than that of the previous study [14]. In addition, the findings of this study are also consistent with those of a previous study [42] in which limited apps were found...
to be available for patient dietary self-management (8.7% vs 9.0%). Despite different evaluation aspects, we agreed with the study conducted by Lee et al [61], which found that commercial renal apps had limited capability to support renal patient self-management.

Conclusions

Although most of the available commercial dialysis diet apps are free and easy to use, they are subject to a possible language barrier, theory deficiency, and a lack of credibility, food databases, and tailored education. Thus, they might have limited potential to promote user engagement and patient dietary self-management. Further research efforts are needed to develop a theory- and evidence-based dialysis diet app equipped with desirable features to promote dietary self-management in the dialysis population.

Acknowledgments

J-HL and ZAMD designed the study, interpreted the data, and drafted the manuscript. C-K-ML and II evaluated the apps. JS, MHMZ, and NFZ provided intellectual content and critical review of the manuscript. All authors have reviewed, commented on, and approved the manuscript. This study was supported by an internal grant (Inisiatif Putra Muda, GP-IPM/2018/9615200) from the Universiti Putra Malaysia. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also acknowledge American Journal Experts (AJE) for proofreading this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The valuable features with descriptions adopted from Mendiola et al, 2015.

[DOCX File, 19 KB - mhealth_v8i6e13808_app1.docx]

Multimedia Appendix 2

The definitions of user interaction for health-behavior theory adopted from Davis et al, 2016.

[DOCX File, 13 KB - mhealth_v8i6e13808_app2.docx]

Multimedia Appendix 3

Scores of the evaluated dialysis diet apps (N=22).

[DOCX File, 21 KB - mhealth_v8i6e13808_app3.docx]

Multimedia Appendix 4

The presence of valuable features in evaluated dialysis diet apps from Google Play and the Apple App Store (N=22).

[DOCX File, 22 KB - mhealth_v8i6e13808_app4.docx]

Multimedia Appendix 5

The presence of health-behavior theory constructs in evaluated renal diet apps for dialysis from Google Play and the Apple App Store (N=22).

[DOCX File, 20 KB - mhealth_v8i6e13808_app5.docx]

References


Abbreviations

AJE: American Journal Experts
CKD: chronic kidney disease
DIMA: Dietary Intake Monitoring Application
ICC: intraclass correlation coefficient
MARS: Mobile Application Rating Scale
mHealth: mobile health

©Jun-Hao Lim, Cordelia-Kheng-May Lim, Imliya Ibrahim, Jazlina Syahrul, Mohd Hazli Mohamed Zabil, Nor Fadhлина Zakaria, Zulfiti Azuan Mat Daud. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 01.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Technologies for Opioid Use Disorder Management: Mobile App Search and Scoping Review

Joseph Nuamah¹, BSc, MSc, PhD; Ranjana Mehta¹, BEng, MEng, PhD; Farzan Sasangohar¹,², BA, BCS, MASc, SM, PhD

¹Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States
²Center for Outcomes Research, Houston Methodist Hospital, Houston, TX, United States

Corresponding Author:
Farzan Sasangohar, BA, BCS, MASc, SM, PhD
Department of Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX, 77843
United States
Phone: 1 9794582337
Email: sasangohar@tamu.edu

Abstract

Background: Advances in technology engender the investigation of technological solutions to opioid use disorder (OUD). However, in comparison to chronic disease management, the application of mobile health (mHealth) to OUD has been limited.

Objective: The overarching aim of our research was to design OUD management technologies that utilize wearable sensors to provide continuous monitoring capabilities. The objectives of this study were to (1) document the currently available opioid-related mHealth apps, (2) review past and existing technology solutions that address OUD, and (3) discuss opportunities for technological withdrawal management solutions.

Methods: We used a two-phase parallel search approach: (1) an app search to determine the availability of opioid-related mHealth apps and (2) a scoping review of relevant literature to identify relevant technologies and mHealth apps used to address OUD.

Results: The app search revealed a steady rise in app development, with most apps being clinician-facing. Most of the apps were designed to aid in opioid dose conversion. Despite the availability of these apps, the scoping review found no study that investigated the efficacy of mHealth apps to address OUD.

Conclusions: Our findings highlight a general gap in technological solutions of OUD management and the potential for mHealth apps and wearable sensors to address OUD.

(JMIR Mhealth Uhealth 2020;8(6):e15752) doi:10.2196/15752

KEYWORDS
mHealth; apps; wearable sensors; substance abuse disorder; mobile phone

Introduction

Background
On average, 5 people in the United States die every hour from an opioid overdose [1]. In 2017 alone, over 70,000 deaths occurred due to drug overdose [2]. This problematic pattern of opioid use often referred to as opioid use disorder (OUD), is considered a public health emergency [1,3] with significant negative impacts on health care [4,5] and criminal justice costs [6]. Misuse of opioids can occur among patients who are initially exposed to opioids in the perioperative period—periods immediately before, during, and after a surgical operation—or through a prescription for the treatment of acute or chronic pain [7]. In addition, opioids attract illegal users and individuals who profit by selling them unlawfully [8]. Such illegitimate use of prescription opioids has exacerbated the increase in OUDs [9-11].

Treatment exists for OUD, comprising pharmacotherapy and behavioral therapies [12,13]. Opioid-dependent users may experience challenging and often severe withdrawal symptoms, including restlessness, muscle aches, and depression, when they abruptly discontinue or reduce opioid intake [14]. Irrespective
of the OUD treatment path, opioid withdrawal management, which includes regularly monitoring patients for symptoms, is the crucial first step after opioid use cessation or dose reduction [1]. A review of opioid withdrawal monitoring methods [15] revealed that the current method of assessing opioid withdrawal using various scales (tools to monitor and rate common signs and symptoms of withdrawal) is self-reported, requires frequent observations, may suffer from recall bias (a study by Infante-Rivard and Jacques [16] gives more details)—unintentional or intentional underreporting of information by respondents—and is ineffective outside of clinic or research environments. Moreover, opioid withdrawal scales differ with regard to the number of scale items and rating criteria. Although technologies such as electronic prescription systems for controlled substances [17], medication history repositories, exchange of clinical records, and clinical direct messaging [8] have been proposed as useful methods to address opioid management, an opioid monitoring method that noninvasively and continuously monitors patients’ symptoms as they occur in real time would provide several distinct advantages over the existing methods [18].

Advancements in technology have allowed the continuous monitoring of diseases outside of clinical settings. Mobile health (mHealth), one such advancement, involves the use of mobile devices to collect health data, monitor signs and symptoms, deliver remote care, and/or educate patients [19]. mHealth interventions allow medical content to be delivered anytime and anywhere to patients [20]. mHealth apps have been used to manage chronic diseases, including monitoring and managing day-to-day symptoms of sickle cell disease [21,22]; monitoring patients undergoing cardiac rehabilitation [23]; monitoring blood pressure measurements to control hypertension [24]; monitoring blood glucose, blood pressure, and physical activity to prevent metabolic syndrome [25]; and monitoring patients with chronic obstructive pulmonary disease [26] (a study by Hamine et al [27] gives a systematic review of mHealth apps for chronic disease management). However, in comparison to chronic disease management, the application of mHealth to OUD has been limited. Digital health technologies, including mHealth apps, have the potential to play a unique role in tackling OUD. These include enabling care providers to create digital profiles of patients to provide personalized care regardless of time and place, monitoring patients’ vital trends and issuing alerts to them or their caregivers, and providing insights into what triggers patients’ behaviors.

Objectives

Inspired by this gap, the overarching aim of our research was to design OUD management technologies that utilize wearable sensors to provide continuous monitoring capabilities. In particular, this research addresses the missed opportunity in monitoring withdrawal symptoms, given their acute nature, salient physiological correlates, and their importance to long-term sobriety. As the first step in investigating novel technological solutions for remote monitoring and management of OUD and, in particular, withdrawal symptoms, we investigated the availability and evidence to support the efficacy of current mHealth and wearable sensor solutions for OUD. The objectives of this paper were to (1) document the currently available opioid-related mHealth apps, (2) review the past and existing technology solutions that address OUD, and (3) to discuss opportunities for technological withdrawal management solutions. To the best of our knowledge, no such review or landscape analysis of technologies that address OUD has been conducted to date.

Methods

Overview

A two-phase parallel search approach was used, which involved an app search to determine the availability of opioid-related mHealth apps, and a scoping review of relevant literature was undertaken to identify relevant technologies and mHealth apps used to address OUD. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews guidelines were used [28].

Mobile Health App Search Method

A search was conducted on the Apple App Store and Google Play for apps published until May 10, 2019, using a combination of search terms that included “opioid”, “opiate”, “substance use disorder”, “technology”, OR “addiction”. The inclusion criteria were as follows: relevance to opioid, opioid prescription, opioid training, opioid monitoring, opioid overdose, opioid addiction support, or substance use disorder (SUD), including opioids. Apps that used a non-English language, apps that solely addressed SUD but not specific to opioids, and apps that required a web browser to use were excluded.

Overall, 2 reviewers independently applied the inclusion and exclusion criteria and identified the final set of apps for review. For each app, the reviewers independently extracted the following: app name, app description, year published, publisher or seller, download estimate, rating, and price. The reviewers transferred the extracted data to a detailed Excel spreadsheet. Then, the reviewers coded the apps for the operating system, that is, Android operating system (henceforth Android) and iPhone operating system (henceforth iOS); clinical focus (opioid-specific or SUD including opioid); audience (patients, clinicians, or anyone); and function (medication-assisted treatment, education, prescription, professional support, peer support, withdrawal support, and patient monitoring; Table 1). Each app was assigned to one primary audience and clinical focus; however, each app could be categorized under more than one app function. Disagreements regarding exclusion/inclusion and coding of the apps were discussed with a third reviewer, and agreement was reached through discussion.

Scoping Review Method

PubMed, Excerpta Medica Database (EMBASE), and Google Scholar were searched for articles published from their inception until May 10, 2019, using a combination of search terms: (“wearable” OR “sensors” OR “technology” OR “mHealth” OR “app” OR “mobile”) AND (“opioid use disorder” OR “opioid” OR “opiate”). Studies were included if they (1) were in English, (2) were peer reviewed, and (3) employed wearable sensors and/or mHealth. Animal studies and studies that did not include opioids were excluded.
The selection of articles was conducted in 2 stages. In the first stage, 2 reviewers independently reviewed titles and abstracts against the inclusion and exclusion criteria using a web-based tool for systematic and scoping reviews called Rayyan [29]. The decision to fully review an article was made when both reviewers agreed to include the abstract. The reviewers resolved disagreements regarding article eligibility by discussing with a third reviewer. In the second stage, full-text articles were reviewed to determine eligibility. Furthermore, backward and forward reference searches were conducted on all full-text articles that met the study selection criteria. 

**Table 1.** Taxonomy used for mobile health app coding.

<table>
<thead>
<tr>
<th>Code and category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Audience</strong></td>
<td></td>
</tr>
<tr>
<td>Patient-facing</td>
<td>App supports patient interactions and engagement</td>
</tr>
<tr>
<td>Clinician-facing</td>
<td>App assists physician decision making</td>
</tr>
<tr>
<td>Anyone</td>
<td>App that is designed for general public, including patients and caregivers</td>
</tr>
<tr>
<td><strong>Clinical focus</strong></td>
<td></td>
</tr>
<tr>
<td>Opioid-specific</td>
<td>App related to only opioids</td>
</tr>
<tr>
<td>Substance use disorder</td>
<td>App related to substances, including opioids</td>
</tr>
<tr>
<td><strong>App function</strong></td>
<td></td>
</tr>
<tr>
<td>Medication-assisted treatment</td>
<td>App supports medication-assisted treatment of opioid use disorder</td>
</tr>
<tr>
<td>Education</td>
<td>App provides educational information</td>
</tr>
<tr>
<td>Conversion</td>
<td>App helps generate equivalent doses of various oral and intravenous opioids</td>
</tr>
<tr>
<td>Professional support</td>
<td>App provides connections to outside professional support, eg, sends a message through the app to seek immediate emergency assistance, finds services and resources that are available nearby</td>
</tr>
<tr>
<td>Peer support</td>
<td>App provides connections to peer support, including individuals undergoing rehabilitation</td>
</tr>
<tr>
<td>Withdrawal support</td>
<td>App supports patients as they go through withdrawal with, eg, reminders, supportive messages, symptom library</td>
</tr>
<tr>
<td>Patient monitoring</td>
<td>App prompts patients to self-evaluate and submit regular personal assessments directly for the purpose of tracking progress and patterns of behavior</td>
</tr>
</tbody>
</table>

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram showing the process of searching and selecting studies included in the review. EMBASE: Excerpta Medica Database.
Furthermore, 2 reviewers independently read the full text of each article identified for inclusion in the review to extract pertinent data using a data extraction form. From each article, the reviewers independently extracted the following: technologies used, physiological parameters, functions, research methods employed, and study findings. The reviewers transferred the extracted data to a detailed Excel spreadsheet. The technologies used were further organized into ecological momentary assessment (EMA), GPS information, wearable sensors, machine learning, and biomedical devices.

**Results**

### Mobile Health App Search Results

The search yielded a total of 72 apps. Of the 72 apps, 62 apps (86%) were available for download at no cost. The remaining 10 (10/72, 12%), all clinician-facing apps, had prices ranging from US $0.99 to US $9.99. Figure 2 shows the number of apps that were made available from January 2009 to May 10, 2019, for both operating systems. Table 2 shows the apps categorized by the audience and the operating system. Clinician-facing apps were most frequently available (31/72, 43%), followed by apps that could be used by patients, caregivers, or the general public (23/72, 32%). As shown in Table 3, most of the available apps were opioid-specific (62/72, 86%).

![Graph showing the number of apps published from January 2009 to May 10, 2019. iOS: iPhone operating system.](image)

**Table 2.** Apps categorized by audience and operating system.

<table>
<thead>
<tr>
<th>Operating system</th>
<th>Apps categorized by audience, n (%)</th>
<th>Total apps, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient-facing</td>
<td>Clinician-facing</td>
</tr>
<tr>
<td>Android only</td>
<td>3 (23)</td>
<td>8 (61)</td>
</tr>
<tr>
<td>iOS* only</td>
<td>1 (5)</td>
<td>14 (82)</td>
</tr>
<tr>
<td>Both Android and iOS</td>
<td>14 (33)</td>
<td>9 (21)</td>
</tr>
<tr>
<td>Total</td>
<td>18 (25)</td>
<td>31 (43)</td>
</tr>
</tbody>
</table>

* iOS: iPhone operating system.

**Table 3.** Apps categorized by clinical focus and operating system.

<table>
<thead>
<tr>
<th>Operating system</th>
<th>Apps categorized by clinical focus, n (%)</th>
<th>Total apps, n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Opioid-specific</td>
<td>Substance use disorder</td>
</tr>
<tr>
<td>Android only</td>
<td>11 (84)</td>
<td>2 (15)</td>
</tr>
<tr>
<td>iOS* only</td>
<td>15 (88)</td>
<td>2 (11)</td>
</tr>
<tr>
<td>Both Android and iOS</td>
<td>36 (85)</td>
<td>6 (14)</td>
</tr>
<tr>
<td>Total</td>
<td>62 (86)</td>
<td>10 (13)</td>
</tr>
</tbody>
</table>

* iOS: iPhone operating system.

Furthermore, apps were analyzed for utilities (Table 4). Although most apps provided opioid conversion support (25/72, 35%) or educational content (21/72, 29%), only 2 opioid-specific apps (2/62, 3%), namely, FlexDek for medication-assisted treatment (MAT) and MATx by Substance Abuse and Mental Health Services Administration (SAMHSA), were designed to support medication-assisted treatment and 4 apps (4/72, 5%) provided support for patient monitoring.
Most apps (25/72, 35%), all clinician-facing and opioid-specific, were developed to convert from one opioid to another. These were also the most downloaded apps (Table 5). For example, Opioid Converter (Figure 3), the app with the highest number of downloads, is a free app supported by Emory University and is designed to aid with opioid dose conversions. The app has a slider that allows for adjustments to be made for incomplete cross-tolerance. The opioids covered include buprenorphine, butorphanol, codeine, fentanyl, hydrocodone, morphine, and oxycodone.

**Table 4.** App tallies for different function categories (utilities are not mutually exclusive).

<table>
<thead>
<tr>
<th>Clinical focus</th>
<th>App categorized as per their functionality, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medication-assisted treatment</td>
</tr>
<tr>
<td>Opioid-specific (n=62)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Substance use disorder (n=10)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Total (n=72)</td>
<td>3 (4)</td>
</tr>
</tbody>
</table>

**Table 5.** Most downloaded Android apps.

<table>
<thead>
<tr>
<th>App name</th>
<th>Year published</th>
<th>Rating (out of 5)</th>
<th>Reviews, n</th>
<th>Estimated downloads, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opioid Converter</td>
<td>2011</td>
<td>4.0</td>
<td>170</td>
<td>50,000+</td>
</tr>
<tr>
<td>Orthodose</td>
<td>2013</td>
<td>4.6</td>
<td>56</td>
<td>10,000+</td>
</tr>
<tr>
<td>Opioid Calculator</td>
<td>2016</td>
<td>4.0</td>
<td>34</td>
<td>10,000+</td>
</tr>
<tr>
<td>CDC® Opioid Guideline</td>
<td>2016</td>
<td>2.8</td>
<td>17</td>
<td>10,000+</td>
</tr>
<tr>
<td>Painkiller Calculator</td>
<td>2014</td>
<td>4.2</td>
<td>21</td>
<td>5000+</td>
</tr>
<tr>
<td>FEND by Preventum</td>
<td>2018</td>
<td>4.2</td>
<td>32</td>
<td>5000+</td>
</tr>
</tbody>
</table>

*CDC: Centers for Disease Control and Prevention.

**Figure 3.** Screenshots of Opioid Converter app showing the main interface (left), selection of an opioid (center), and 25 mg oxycodone adjusted at 40% for incomplete cross-tolerance (right).

Overall, 9 of the 72 apps (9/72, 12%) were designed to provide professional support, including connecting users with a network of service providers and finding naloxone carriers in an overdose emergency. Furthermore, 6 of the 72 apps (6/72, 8%) were designed to provide peer support in the form of reminders, supportive messages, and symptom library. In addition, 4 out of 72 apps (4/72, 6%) were designed to provide patient monitoring by prompting patients to self-evaluate and submit...
regular personal assessments directly for the purpose of tracking progress and patterns of behavior. Overall 2 out of 72 apps (2/72, 3%) were categorized as other. One of these apps, Diagnosis, Intractability, Risk, and Efficacy (DIRE), is designed for clinicians to use DIRE as a tool [30] in their decision-making process when considering prescribing opioids. The DIRE tool allows clinicians to rate 7 factors (diagnosis, intractability, psychological risk, chemical health risk, reliability risk, social support risk, and efficacy) on a scale of 1 to 3, with 1 being the least favorable case for prescribing opioids and 3 being the most favorable case for prescribing opioids. The total score (i.e., the sum of the ratings) is used to determine a patient’s suitability for opioid maintenance analgesia. The other app is THRIVEE, a virtual platform system designed to help patients overcome addiction. THRIVEE delivers virtual MAT to addicts, including opioid users. It utilizes virtual telemedicine sessions such as video conferencing between patients and providers to leverage proven clinical practices.

The total number of downloads was used as a measure of app prevalence. Although the download statistics were not available for iOS apps, the statistics for Android apps varied from as low as 5+ downloads to as high as 50,000+ downloads (Figure 4). Table 5 shows the 6 most downloaded Android apps and their respective ratings.

**Figure 4.** Timeline of the apps showing the year each app was first published (on the horizontal axis) versus the estimated number of downloads from the date the app was published to the search date (on the vertical axis). Timeline for most downloaded Android apps showing the number of downloads from January 2010 to May 10, 2019. Download statistics are not available for iPhone operating system–based apps. FEND: Full Energy No Drugs. CDC: Centers for Disease Control and Prevention. OARS: Opioid Addiction Recovery Support. MAT: Medication Assisted Treatment.

**Focused Review Results**

Our initial search yielded 6459 articles. These were exported to the Zotero reference management software, where 842 duplicates were removed. Title and abstract screening resulted in the exclusion of 5593 articles. The remaining 24 articles were fully reviewed. Of these 24 articles, 18 met the inclusion criteria and were included in the final review.

Our search yielded 18 papers that documented relevant technologies used to address OUD. Of the 18 studies, 9 (50%) were laboratory-based studies, 8 (44%) were field studies, and 1 (6%) was a clinical trial. We did not find studies that employed mHealth apps to address OUD. Table 6 presents a summary of the technologies identified in the scoping review.
<table>
<thead>
<tr>
<th>Article</th>
<th>Technologies</th>
<th>Physiological parameters</th>
<th>Utility</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epstein et al [31]</td>
<td>PDA(^a) (Palm Zire, PZ21) and diary software</td>
<td>N/A(^b)</td>
<td>Monitoring</td>
<td>5 random prompts per day (5 weeks) and 2 random prompts per day (20 weeks)</td>
</tr>
<tr>
<td>Boyer et al [32]</td>
<td>Smartphones, wearable sensors, and machine learning</td>
<td>EDA(^c), acceleration, skin temperature, and heart rate</td>
<td>Real-time detection of drug craving and interventions</td>
<td>Self-annotation of physiological changes and machine learning</td>
</tr>
<tr>
<td>Epstein and Preston [33]</td>
<td>PDA (Palm Zire, Palm Zire 21) and diary software</td>
<td>N/A</td>
<td>Momentary ratings of stress in outpatients at work</td>
<td>5 random prompts per day (5 weeks) and 2 random prompts per day (20 weeks)</td>
</tr>
<tr>
<td>Kennedy et al [34]</td>
<td>PDA (Palm Zire, PZ21) and diary software</td>
<td>N/A</td>
<td>Gender-based treatment strategies</td>
<td>Random prompts (2-5 per day) for location, activities, and companions</td>
</tr>
<tr>
<td>Epstein et al [35]</td>
<td>PDA (PalmPilot) and GPS (BT-Q1000X)</td>
<td>N/A</td>
<td>Real-time monitoring of mood, stress, and drug craving</td>
<td>Time-stamped GPS data and EMA(^d) ratings of mood, stress, and drug craving</td>
</tr>
<tr>
<td>Kennedy et al [36]</td>
<td>Biosensor (AutoSense) and smartphone</td>
<td>Heart rate</td>
<td>Continuous monitoring of heart rate</td>
<td>Wireless heart rate sensor data and self-reports</td>
</tr>
<tr>
<td>Carreiro et al [37]</td>
<td>Biosensor (Q sensor)</td>
<td>EDA, skin temperature, and acceleration</td>
<td>Real-time detection of drug use</td>
<td>Continuous monitoring of EDA, skin temperature, and acceleration</td>
</tr>
<tr>
<td>Linas et al [38]</td>
<td>PharmChek drugs of abuse patches, Palm Z22, and smartphone</td>
<td>Sweat patches detect traces of cocaine or heroin secreted in sweat during the period they are worn</td>
<td>Agreement of EMA methods with other methods (ie, biological and ACASI(^e) of assessing drug use</td>
<td>Palm Z22 PDA (3 trials) and Motorola Droid X2 phone (1 trial), self-reports of heroin or cocaine, sweat patches (weekly), and ACASI (weekly)</td>
</tr>
<tr>
<td>Mennis et al [39]</td>
<td>Smartphone and GPS</td>
<td>N/A</td>
<td>Integration of GPS information with EMA to study neighborhood effects on opioid use disorder</td>
<td>Combined GPS information with EMA to find association among neighborhood disadvantage, perceived stress, perceived safety, and substance use; generalized estimated equations for analysis</td>
</tr>
<tr>
<td>Sarker et al [40]</td>
<td>Biosensor, smartphone, GPS, and machine learning</td>
<td>ECG(^f) and inspiratory to expiratory ratio</td>
<td>Time series health data to determine the timing of interventions and links to prevent drug craving and relapse</td>
<td>Smartphone initiated 32-item EMA (random); modeling R-R intervals and heart rate variability from ECG data</td>
</tr>
<tr>
<td>Carreiro et al [18]</td>
<td>Biosensor (Q sensor)</td>
<td>EDA, skin temperature, and acceleration</td>
<td>Biosensors may be used in drug addiction treatment and pain management</td>
<td>Hilbert transform analyses combined with paired t-tests to compare biosensor data</td>
</tr>
<tr>
<td>Wang et al [41]</td>
<td>Biosensor (Q sensor), urine drug screens, and patient self-report of substance use</td>
<td>EDA, skin temperature, and acceleration</td>
<td>Detect and set up thresholds of parameters in real-time drug use event detection for wearable biosensor data streams</td>
<td>Sliding window technique to process data stream and distance-based outlier algorithm to detect substance use events</td>
</tr>
<tr>
<td>Chintha et al [42]</td>
<td>Biosensor (Empatica E4)</td>
<td>Skin temperature, acceleration, and heart rate</td>
<td>Identify physiologic change that marks wearing off of naloxone effect</td>
<td>90-min postnaloxone time point evaluated with Hilbert transform</td>
</tr>
<tr>
<td>Kowalczyk et al [43]</td>
<td>PalmOne Zire 21, Palm Tungsten E2, or HTC TyTN II smartphone</td>
<td>N/A</td>
<td>Investigate the relationship between opioid use and craving and affect</td>
<td>Mobile devices used to rate craving 4 times randomly each day</td>
</tr>
<tr>
<td>Mahmud et al [44]</td>
<td>Biosensor (Q sensor) and machine learning</td>
<td>EDA and skin temperature</td>
<td>Automatic detection of opioid intake and classification of pre- and postopioid health conditions</td>
<td>Time and frequency domain feature analysis; decision tree, k-nearest neighbors, and extreme gradient boosting classifiers</td>
</tr>
<tr>
<td>Moran et al [45]</td>
<td>Smartphone</td>
<td>N/A</td>
<td>Gender differences in the influence of stress on opioid use and craving</td>
<td>Entry was initiated, and causes, context, stress, and craving severity were rated each time the participant felt more stressed than usual</td>
</tr>
</tbody>
</table>
Perceived safety, and greater substance use, suggesting that GPS neighborhood disadvantage, higher perceived stress, lower drug craving, suggesting that mood, stress, and drug craving in a geographical context [35] and to examine the relationship between daily hassles and stressful events in opioid-dependent men and women [46]. Epstein and Preston [33] found opioid-dependent outpatients to be less stressed at the workplace than elsewhere, demonstrating the utility of EMA to rate stress in outpatients. Kennedy et al [34] found that males and females with SUD differ in their daily functioning during addiction treatment, highlighting the need to develop gender-based treatment strategies. Similarly, Moran et al [45] found that stress-induced craving differs between opioid-dependent men and women, suggesting that gender-based tailoring of treatment should consider individual differences. Kowalczyk et al [43] found that cravings increased when the participants were using opioids, indicating the utility of EMA to investigate the relationship between opioid use and craving. Overall, EMA has shown promise in enabling the measurement of momentary experiences and states of cravings and misuse in natural settings.

**GPS Information**

Overall, 2 studies (2/18, 11%), both field-based, combined EMA with GPS location information to monitor the real-time mood, stress, and drug craving in a geographical context [35] and to study neighborhood effects on substance use [49]. EMA provided the participants’ momentary experience, whereas GPS provided the participants’ location during those experiences. Epstein et al [35] found a negative association among environmental disorders (defined as a lack of order and social control within the neighborhood) [49] and mood, stress, and drug craving, suggesting that mood, stress, and drug craving can be monitored in real time in a geographical context. Mennis et al [39] found a significant positive association among neighborhood disadvantage, higher perceived stress, lower perceived safety, and greater substance use, suggesting that GPS information can be combined with EMA to study neighborhood effects on substance use.

**Wearable Sensors**

The advances in wearable technologies have enhanced the ability of researchers to monitor physiological changes associated with opioid intake and/or drug craving. Overall, 8 out of the 18 studies (8/18, 44%) employed wearable sensors. Of these 8 studies, 3 (37%) studies [32,36,40] combined EMA and wearable sensors to detect drug cravings [32,36], deliver personalized prevention interventions [32], and determine stress episodes among opioid users [40]. Kennedy et al [36] reported higher heart rates when participants reported craving compared with when they reported no craving, suggesting the potential efficacy of using heart rate data for continuous monitoring of craving. The iHeal system [32]—a system architecture intended to provide personalized interventions—combines EMA, wearable sensors, and a deep belief network model to detect drug cravings and deliver personalized drug prevention interventions. However, this study did not implement their iHeal system.

The remaining 5 (5/8, 63%) studies [18,37,41,42,44] used wearable biosensors for real-time detection of opioid use [37,41] to detect physiological changes associated with opioid use [18], evaluate physiological changes associated with the wearing off of naloxone [42], and automatically detect opioid intake [44]. Studies using Q sensors, worn on the participants’ wrists, have found that an increase in electrodermal activity (EDA) is associated with opioid use [41], accurately detected substance use events within 30 min [41] and significant within-subjects increase in skin temperature and decrease in locomotion immediately after opioid administration [18]. Carreiro et al [18] found that physiological changes varied among subjects with the levels of opioid use—heavy opioid users showed a greater decrease in fidgeting movements than nonheavy opioid users. Chinthu et al [42] used an E4 device (Empatica) worn on the participants’ wrists and found that heat rate and skin temperature differed significantly between before and after naloxone administration. Finally, Linas et al [38] combined EMA and wearable sweat patches, PharmChek Drugs of Abuse Patches (PharmChem Inc), to concurrently collect momentary data and sweat in the field from 109 adults with recent opioid intake.
use and found moderate-to-good agreement of EMA to sweat patches and self-report methods in capturing drug use events.

**Machine Learning**

Overall, 4 studies (4/18, 22%) used machine learning techniques to analyze and predict opioid use. Furthermore, 3 (3/18, 17%) of these studies [40,41,44] predicted opioid intake. The remaining study [39] developed a model to provide personalized interventions. Sarker et al [40] combined EMA, location information; the cStress model (reported on in a study by Housepian et al [50]), which uses electrocardiogram (ECG) and respiration data; and the moving average convergence divergence method to predict stress episodes associated with opioid intake. Their model predicted stress episodes with an accuracy of 94.8% and kappa of 0.444. Wang et al [41] used a sliding window technique to process streams of EDA, skin temperature, and acceleration data collected from a wrist-worn Q sensor and a distance-based outlier algorithm to detect substance use events. Their model accurately detected substance use events within 30 min. Using 2 parameters, movement in the z-axis and skin temperature collected from wrist-worn Q sensors, Mahmud et al [44] compared the ability of 3 classifiers (decision tree, k-nearest neighbors, and extreme gradient boosting) to automatically detect opioid intake, obtaining an accuracy of 99.4% with extreme gradient boosting.

**Biomedical Devices**

Miranda and Taca [47] investigated the effect of an auricular neurostimulation device, BRIDGE, in treating opioid withdrawal symptoms. The device was placed behind the ears of 73 opioid-dependent outpatients for a maximum of 5 days to treat opioid withdrawal symptoms by stimulating the nerves in the brain and spinal cord. A reduction in opioid withdrawal scores, measured with the clinical opioid withdrawal scale, was associated with the use of BRIDGE.

**Discussion**

**Principal Findings**

The goal of the app search in this study was to determine the availability of opioid-related mHealth apps. The search revealed the availability of 72 Android and iOS apps and revealed a steady rise in app development from January 2009 to May 10, 2019, with most apps designed to support clinicians. Our findings suggest that most of the apps have been developed to help clinicians convert from one opioid to another at an equianalgesic dose. Despite the availability of these apps, the scoping review found no study that investigated the efficacy of mHealth apps to address OUD.

**Outcome Evaluation of Mobile Health App Search**

Opioid conversion, a common but challenging clinical practice [51], is required when patients do not respond therapeutically, develop adverse effects to an opioid, or need an alternative route of administration [52]. Prescription error has been identified as a significant risk factor for opioid-related deaths [53]; therefore, opioid conversion apps that run on mobile devices may help improve patient safety [54]. Although these apps are not geared toward OUD, they help primary care providers safely prescribe opioids.

The US Food and Drug Administration (FDA) has the mandate to regulate mHealth apps that meet certain statutory criteria as medical devices. Under the existing FDA regulatory framework, it is difficult to determine whether an mHealth app is a medical device [55]. The FDA has long exempted apps considered low-risk from its approval process [56]. It is unclear how many opioid conversion apps identified in this study have been approved by the FDA. For example, although Pear reSET-O, a prescription app, was first published in 2016, it is only recently that the FDA cleared it as the first software-generated therapeutic intervention for patients with OUD [57]. This app provides cognitive behavioral therapy to patients enrolled in an OUD treatment program. Although most apps identified in this study are free to download, many health care providers and patients may not be aware of their availability. Future studies should investigate such awareness and adoption rates. The factors that influence the adoption of mHealth apps by health professionals include lack of clinical evidence [58], security [59], and inability to integrate apps with other systems [60]. The factors that influence patients’ adoption of mHealth apps include security and privacy concerns [61,62], social contacts [63], and cost of smartphones and data plans [62,64]. Failure to balance system demands of apps with end user needs and resources undermines the adoption of mHealth apps [65]. Conducting content analyses, usability testing, observational studies, and efficacy testing will contribute to the increased adoption of mHealth apps in clinical practice [66].

**Privacy of Mobile Health Apps**

The privacy of mHealth apps, the right for users to know how their information is collected and used, is an issue worthy of discussion. In this study, most apps identified in the search were free to download. For users of these apps, there is a likelihood that their information is passed around to third parties, thereby exposing them to privacy risks [67]. A recent study investigated data sharing practices in the mHealth ecosystem and found that 79% of the sampled apps shared user data with 55 entities, including third parties [68]. This presents a major concern for mHealth users as they do not know how their data will be used and by whom. Furthermore, the aggressive medicolegal system in the United States deters many health care providers from using mHealth apps. Recent studies (eg, study by Hutton et al [69]) have suggested the need for standards that can ensure mHealth app user privacy.

**Outcome Evaluation of Focused Review**

Despite the availability of opioid-related apps, this scoping review, which sought to document relevant technology solutions that address OUD, found no study that employed mHealth apps to address OUD. Most of the studies employed EMA to capture participants’ opioid use patterns as they occurred in real time. Few studies have combined EMA with a range of data types, including physiological changes and location information, to detect opioid intake. These findings highlight a general lack of empirical evidence to support the efficacy of mHealth apps for...
OUD management. However, our findings show the potential for wearable sensors, especially in opioid withdrawal management, to facilitate remote monitoring of the signs and symptoms of OUD.

Opioid withdrawal management, which includes regular monitoring of patients for symptoms, is the crucial first step after opioid use cessation or dose reduction [1]. The relapse rates during inpatient treatment of opioid dependence indicate that as many as 91% of those in recovery experience an opiate relapse, 59% of whom relapse within the first week of sobriety, and 80% within a month after discharge from a detox program [70]. The results from this scoping review revealed that most of the studies employed EMA or combined EMA with a range of data types to detect opioid usage patterns. These studies focused on opioid intake and usage patterns. Only one study [47] focused on developing technology to help treat opioid withdrawal symptoms. Indeed, the BRIDGE device used in that study is the first of its kind approved by the FDA. It is crucial that technology solutions be provided not only to help health care professionals monitor and manage patients’ withdrawal symptoms but also to help the patients themselves as they go through withdrawal.

Gaps Identified in Outcome Evaluation of Technologies for Opioid Use Disorder Management

From the results of this study, it is evident that there is a gap in the technologies available to manage opioid withdrawal. Advances in wearable and machine learning technologies have enhanced the ability of researchers to monitor physiological changes associated with opioid intake and/or drug craving [40,41,44]. In the same vein, wearable sensors can be employed to detect temporal and spectral patterns of physiological responses associated with opioid withdrawal symptoms. For example, joint/muscle aches lead to elevated heart rate [71], which can be measured with a wearable ECG; anxiety leads to elevated heart rate [72] and change in skin conductance [73], which can be measured using wearable ECG and EDA sensors, respectively; and cutis anserine, defined as goosebumps, leads to a change in skin conductance [74], which can be measured with a wearable EDA sensor. Machine learning–based pattern detection algorithms may be used to explicitly detect and characterize specific features obtained from wearable sensor configurations and existing contextual information. This can provide real-time feedback to health care providers to facilitate interventions.

Limitations
There are some limitations in the study that warrant discussion. First, the search may not be collectively exhaustive because of the limitations of the scoping review. The scoping review utilized relatively fewer, albeit relevant, number of search terms and databases to identify potentially eligible studies. Despite this limitation, we found saturation in the technologies used to address OUD, evidenced by the lack of additional results from the 19-article–based bibliographic secondary search. Second, the availability of information about app downloads was limited to Android apps only. However, the data presented in this study are relevant, given that Android has overtaken iOS as the number 1 operating system for mHealth apps [75]. Third, although the app rating information is reported, it is difficult to determine how many of the ratings were legitimately written by people who used the apps. In addition, we were unable to determine how the apps were rated. Owing to this lack of information, this study did not include information on the quality of the apps. Furthermore, we did not focus on capturing the apps’ effectiveness. Given the proliferation of mHealth apps and technologies made available to target OUD, future studies should aim to investigate the quality and effectiveness of these apps on OUD management. Finally, developers may be reluctant to publish research on their apps for intellectual property reasons (if they have any); many of their results/algorithms may be considered proprietary.

Conclusions
This study showed the availability of opioid-relevant mHealth apps, most of which are opioid conversion apps. Despite the availability of these apps, the scoping review found no study that employed mHealth apps to address OUD. Most studies employed EMA to capture the participants’ opioid usage patterns as they occurred in real time. Few studies have combined EMA with a range of data types, including physiological changes and location information, to detect opioid intake. Our findings highlight the gap in technologies and the potential for using wearable sensors, especially in opioid withdrawal management, to address OUD.

Acknowledgments
This independent work was supported by the Texas A&M Triads for Transformation. The authors would like to thank Dr. Ethan Larsen for his assistance during the app review process.

Authors’ Contributions
RM and FS conceptualized the study, JN conducted and analyzed the reviews and drafted the initial manuscript, and RM and FS interpreted the review outcomes and refined the manuscript.

Conflicts of Interest
None declared.

References
None declared.


Abbreviations

- DIRE: Diagnosis, Intractability, Risk, and Efficacy
- ECG: electrocardiogram
- EDA: electrodermal activity
- EMA: ecological momentary assessment
- FDA: Food and Drug Administration
- iOS: iPhone operating system
- mHealth: mobile health
- OUD: opioid use disorder
- SUD: substance use disorder

©Joseph Nuamah, Ranjana Mehta, Farzan Sasangohar. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 05.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
The Quality of Mobile Apps Used for the Identification of Pressure Ulcers in Adults: Systematic Survey and Review of Apps in App Stores

Janine Koepp1,2*, PhD; Miriam Viviane Baron2,3*, MSc; Paulo Ricardo Hernandez Martins4*, BA; Cristine Brandenburg5,6*, MSc; Ariane Tieko Frare Kira2*, MD; Vanessa Devens Trindade2,6*, MD; Luis Manuel Ley Dominguez4*, BA; Marcelo Carneiro1*, MD, PhD; Rejane Frozza1*, PhD; Lia Gonçalves Possuelo1*, PhD; Marcus Vinicius De Mello Pinto7*, PhD; Liane Mahlmann Kipper1*, PhD; Bartira Ercília Pinheiro da Costa4*, PhD

1University of Santa Cruz do Sul, Santa Cruz do Sul, Brazil
2Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil
3Coordination of Improvement of Higher Education Personnel – Brazil, Brasilia, Brazil
4School of Medicine, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil
5Federal University of Ceará, Fortaleza, Brazil
6National Council for Scientific and Technological Development – Brazil, Brasilia, Brazil
7Celulare Institute, Rio de Janeiro, Brazil
*all authors contributed equally

Corresponding Author:
Janine Koepp, PhD
University of Santa Cruz do Sul
Avenida Independência, 2293 - Block 35
Santa Cruz do Sul, 96816-501
Brazil
Phone: 55 51 99806 2275
Email: janinekoepp@yahoo.com.br

Abstract

Background: The increasing global use of smartphones has contributed to the growing use of apps for various health conditions, showing promising results. Through mobile apps, it is possible to perform chronological and iconographic follow-up of wounds, such as pressure ulcers, using a simple and practical tool. However, numerous surveys have pointed out issues related to the functionality, design, safety, and veracity of app information.

Objective: The objective of this study was to perform a systematic review of published studies regarding mobile apps and a systematic survey in app stores looking for apps developed to identify, evaluate, treat, and/or prevent pressure ulcers in adults, and to evaluate those apps based on software quality characteristics.

Methods: This review followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The main bibliographic databases were searched between January 1, 2007 and October 15, 2018, and an app survey was performed in app stores. The selected studies were evaluated according to software quality characteristics by the International Organization for Standardization/International Electrotechnical Commission (ie, ISO/IEC 25010:2011) that involve functionality, efficiency, compatibility, usability, reliability, safety, maintenance, and portability.

Results: The search in databases and web-based app stores returned a total of 2075 studies. After removal of duplicates and screening of titles and abstracts, 48 complete articles were evaluated for eligibility, and among these, six were included for qualitative synthesis.

Conclusions: In this review, it was observed that all studies involved the initial phase of app development or improvement, and therefore, the apps still need to be evaluated using different software quality characteristics, so that in the future, a gold standard can be approached. Therefore, the prescription of an app for the identification, evaluation, treatment, and/or prevention of pressure ulcers in adults is currently limited. However, the evaluated studies provided important insights for future research. It is of utmost importance that future surveys develop apps jointly with users, using collaborative and cocreative processes and assess patients in real-world situations across different service settings, and they should consider different ethnicities, so that apps are useful to
end users, such as patients, family members, health professionals, and students, in the health area. In addition, it is necessary for studies to describe the methodological course of app development in a clear and objective way in order to ensure reproducibility of the study and to offer inputs to allow future research to approach the development of ideal apps that are geared to positively impact the health of end users.

**Trial Registration:** PROSPERO CRD42018114137; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=114137

(JMIR Mhealth Uhealth 2020;8(6):e14266) doi:10.2196/14266

**KEYWORDS**
software; portable app; mobile app; pressure sore; decubitus ulcer; wounds and injuries

## Introduction

### Background

A pressure ulcer involves localized damage to the underlying skin or soft tissues, resulting in localized tissue destruction related to lack of blood flow due to increased external pressure on bone prominence or due to the use of a medical device [1].

Pressure ulcers negatively impact the quality of life of patients, contribute to pain and suffering, prolong hospitalization, increase workload, and increase costs for health systems [2-9]. In addition, the incidence and prevalence rates remain high in different populations and countries. Studies in the intensive care units (ICUs) of hospitals in Brazil showed an incidence of pressure ulcers between 17.2% and 41.0% [10,11]. In the United States, the prevalence of pressure ulcers in the ICU ranges from 8.8% to 12.1%, and in acute care units, it can reach 22% [12].

A pressure ulcer is a wound that is characterized by rapid deterioration of soft tissues and a process of chronification that hinders normal healing. In this way, systematic follow-up of the evolution of the wound by a physician and health team is unavoidable. However, evaluation and follow-up by specialized professionals in loco are not always possible, especially in situations where the patient cannot count on special transportation to a specialized care center or the patient does not have a family member or resident in remote areas. With the emergence of mobile health (mHealth) and the popularity of mobile devices in clinics and hospitals, wound evaluation can now be optimized by allowing an interprofessional team to remotely view, analyze, and monitor wound evolution through apps [13].

A recent study has shown that through an app developed for the management of pressure ulcers, it is possible for a caregiver to show the patient a digital image of a wound on the buttocks, the back region, or under the foot. In this way, the approach brings to patients and families a better understanding of the wound and subsequent compliance with wound treatment guidelines. In addition, app image algorithms can calculate the size of the wound, and additionally, color analysis can aid in the detection of the depth and stage of the pressure ulcer [2], facilitating the monitoring of the evolution of the wound and the choice of correct treatment. Another app developed for pressure ulcer prevention allows users to monitor the pressure of a seat interface in real time, leading to pressure relief maneuvers and allowing the transfer of monitored data for follow-up by a specialist [9]. Advances in this area have shown that apps can assist health professionals in the prevention and treatment of pressure ulcers and facilitate the involvement of family members and patients in their own care, improving management and wound outcomes [2].

Currently, there is an overall increase in app use for various health conditions [14-30]. The convenience in using apps is associated with the many resources available through smartphones. With these facilities, the user may acquire skills and confidence and may adapt quickly to the use of the tool [23]. However, researchers have analyzed apps geared toward the management of various health problems and have discovered several problems with regard to navigability, usability, functionality, design, accuracy, unnecessary resources, lack of free apps, and lack of certification of the quality of the information conveyed, and most apps access personal data on devices. These shortcomings have cast doubts on the applicability and efficacy of mobile apps in various health care sectors [16,18,26,27,29,31,32].

Therefore, knowing if mobile apps used to track pressure ulcers have quality features like functionality, reliability, usability, efficiency, compatibility, security, maintenance, and portability is of extreme relevance to end users. In this context, the research question was elaborated in the format of the acronym PICO (participants, intervention, comparison, outcomes) and was formulated as follows: do mobile apps used by adults to identify, evaluate, treat, and/or prevent pressure ulcers present software quality characteristics?

### Objective

To answer the research question, the aim of this study was to conduct a systematic review of published studies on mobile apps and a systematic survey in app stores for apps developed to identify, evaluate, treat, and/or prevent pressure ulcers in adults, as well as to evaluate apps based on software quality characteristics.

## Methods

### Review Protocol

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for the design of this study, as well as to report the findings of the review [33]. The protocol of this systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO; ID: CRD42018114137).
Search Strategy in Databases and App Stores

A literature search was carried out in collaboration with librarians with experience in systematic reviews. For the search and selection of studies, the following databases were selected: PROSPERO, PubMed, Cochrane Library, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Web of Science, Institute of Electric and Electronic Engineering (IEEE) Xplore Digital library, Compendex (Ei Village 2), Association for Computing Machinery (ACM) Digital Library, Science Direct, Scopus, Scientific Electronic Library Online (SciELO), Literatura Latino-americana e do Caribe em Ciências da Saúde (Latin American and Caribbean Health Sciences Literature database) (LILACS), Google Scholar, and Brazilian Registry of Clinical Trials (ReBEC). The search strategy with key words was initially developed for PubMed (Textbox 1) and later adapted to the other databases according to the syntax required in each database. The online search was carried out from October 6 to November 5, 2018. Additional bibliographies were searched in the references of relevant studies, with contacting of authors and search for gray literature. Additionally, a search was performed for the names of apps about pressure ulcers in web-based stores and subsequently for queries that involved the names of the apps found.

Textbox 1. PubMed search strategy.

```
("nursing"[Subheading] OR "nursing"[All Fields] OR "nursing"[MeSH Terms] OR "patient care team"[MeSH Terms] OR "patient"[All Fields] AND "care"[All Fields] AND "team"[All Fields]) OR "patient care team"[All Fields]) AND ("sensitivity and specificity"[MeSH Terms] OR ("sensitivity"[All Fields] AND "specificity"[All Fields]) OR "sensitivity and specificity"[All Fields]) OR accuracy[All Fields]) AND ("android"[All Fields]) OR (iOS[All Fields]) OR ("mobile applications"[MeSH Terms] OR ("mobile"[All Fields] AND "applications"[All Fields]) OR "mobile app"[All Fields]) OR ("cell phone"[MeSH Terms] OR ("cell"[All Fields] AND "phone"[All Fields]) OR "cell phone"[All Fields]) OR ("smartphone"[MeSH Terms] OR "smartphone"[All Fields]) AND ("pressure ulcer"[MeSH Terms] OR ("pressure"[All Fields] AND "ulcer"[All Fields]) OR "pressure ulcer"[All Fields]) OR ("decubitus"[All Fields] AND "ulcers"[All Fields]) OR "decubitus ulcers"[All Fields])
```

Eligibility Criteria

We considered studies published on the internet in English, Portuguese, or Spanish, dated between January 1, 2007 and October 15, 2018. The search was updated on November 5, 2018. It included original research work limited to humans; studies involving apps on handheld devices having Android, iOS, or other operating systems; studies that resulted in the development of software registration; studies with mobile apps that aimed to identify, evaluate, treat, and/or prevent pressure ulcers; studies involving app users, such as health professionals and information technology professionals, over the age of 18 years; studies that used apps to identify, evaluate, treat, and/or prevent pressure ulcers in individuals over the age of 18 years; and studies performed in care settings, such as university, hospital, and community settings.

The exclusion criteria were as follows: review studies; studies involving mobile apps that, in addition to identifying, evaluating, treating, and/or preventing pressure ulcers, assessed other types of wounds or other health problems; studies involving other electronic information systems; studies involving only electronic forms and electronic medical records; and studies where key information was not available. In the case of duplicate studies, the search considered those with a larger sample size and more information.

Selection Process and Data Extraction

All records were downloaded to Mendeley Desktop Version 1.19.3 and duplicates were removed. All the titles and summaries of the remaining studies were read to identify whether they met the eligibility criteria. Whenever titles and summaries were not sufficient, the full text of the potentially relevant studies was read in its entirety. Two researchers performed the entire process of searching the databases, selecting the studies, and reading the studies independently.

Thereafter, the two authors independently extracted data from the selected studies using a standardized Word table template. The following data were extracted from the studies: author and year of publication, study outline, objective, sample, description of the technology, main results, and characteristics of the quality of the app. Disagreements were resolved by consultation and discussion with a third senior author to reach consensus. The screening and selection of studies are presented in a PRISMA flow diagram (Figure 1).

http://mhealth.jmir.org/2020/6/e14266/
Data Synthesis and Quality Assessment

A narrative and qualitative synthesis was performed. As a summary of the main result metrics, we evaluated any type of result that indicated or measured app quality. This evaluation was based on the standards of the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 25010:2011 called Software Product Quality Requirements and Evaluation (SQuaRE). ISO/IEC 25010:2011 is prescribed for the evaluation of software in production or already developed. This evaluation is based on quality characteristics, such as functionality, efficiency, compatibility, usability, reliability, safety, maintenance, and portability [34].

The original protocol aimed at evaluating any type of study outcome that indicated or measured the accuracy of mobile apps used for the identification, evaluation, treatment, and/or prevention of pressure ulcers, and evaluating the methodological quality of the studies using the Downs and Black instrument (1998) [35]. However, all the included studies involved descriptive searches with apps in the initial development or improvement process (ie, in the preclinical stage of innovation),...
Results

Search Results

A total of 2065 studies were initially identified in the database search, and nine studies were additionally identified through a search of the references of the studies that met the inclusion criteria, through a search of the gray literature, and contact with authors. In addition, the search for apps on pressure ulcers on websites and web-based stores identified 18 apps (Table 1). During the search for names, in order to find published research on the development of apps, two studies were screened; however, only one was retrieved, resulting in a total of 2075 assessed studies. After the removal of duplicate studies and screening of the titles and abstracts, 48 full-text articles were evaluated for eligibility. Of those 48 studies, six were included for qualitative synthesis. The qualitative synthesis of these six studies selected for this review can be found in Table 2. It should be noted that the distribution of the studies by year of publication shows that these six studies were published within the last 10 years.
Table 1. List of apps found on websites and in web-based stores.

<table>
<thead>
<tr>
<th>App</th>
<th>Description</th>
<th>Capture photos of the wound (yes/no)</th>
<th>Research involving the app (yes/no)</th>
<th>Developer</th>
<th>Platform</th>
<th>Device</th>
<th>Language</th>
<th>Paid or free</th>
<th>Country of origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wound Rounds</td>
<td>PU prevention, showing data over the internet</td>
<td>No</td>
<td>No</td>
<td>Tele Medicine Solutions, LLC</td>
<td>iOS</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Paid</td>
<td>USA</td>
</tr>
<tr>
<td>Staging PI</td>
<td>PU classification</td>
<td>No</td>
<td>No</td>
<td>Baylor Scott &amp; White Health</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Free</td>
<td>USA</td>
</tr>
<tr>
<td>Pressure Ulcer</td>
<td>PU prevention</td>
<td>No</td>
<td>No</td>
<td>Patient Data Science, LLC</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Free</td>
<td>USA</td>
</tr>
<tr>
<td>Pressure Ulcer Guide</td>
<td>PU prevention</td>
<td>No</td>
<td>No</td>
<td>Patient Data Science, LLC</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Paid</td>
<td>USA</td>
</tr>
<tr>
<td>Braden Scale 4</td>
<td>PU prevention using Braden scale</td>
<td>No</td>
<td>No</td>
<td>Patient Data Science, LLC</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Paid</td>
<td>USA</td>
</tr>
<tr>
<td>Norton Scale 4</td>
<td>PU prevention using Norton scale</td>
<td>No</td>
<td>No</td>
<td>Patient Data Science, LLC</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Paid</td>
<td>USA</td>
</tr>
<tr>
<td>MOWA</td>
<td>Identifying, assessing, and suggesting PU care</td>
<td>Yes</td>
<td>Yes</td>
<td>Healthpath</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English, Italian, Spanish, French, and Portuguese</td>
<td>Paid</td>
<td>Italy</td>
</tr>
<tr>
<td>PrevenAPP</td>
<td>PU prevention using Braden scale</td>
<td>No</td>
<td>No</td>
<td>Smith &amp; Nephew</td>
<td>INA</td>
<td>INA love for at no longer available</td>
<td>Free</td>
<td>UK</td>
<td>UK</td>
</tr>
<tr>
<td>Riesgo de Úlceras</td>
<td>PU prevention using Braden scale</td>
<td>No</td>
<td>No</td>
<td>Luis Miguel Delgado</td>
<td>Android</td>
<td>Tablet and smartphone</td>
<td>Spanish</td>
<td>Free</td>
<td>Colombia</td>
</tr>
<tr>
<td>Por Presión (Pressure Ulcer Risk)</td>
<td>Provides information and tools for PU prevention, diagnosis, and treatment</td>
<td>No</td>
<td>Yes</td>
<td>ERTAKY</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>Spanish</td>
<td>Free</td>
<td>Spain</td>
</tr>
<tr>
<td>GuíaUPP</td>
<td>Provides information and tools for PU prevention, diagnosis, and treatment</td>
<td>No</td>
<td>Yes</td>
<td>Viacore IT</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>Spanish</td>
<td>Free</td>
<td>Spain</td>
</tr>
<tr>
<td>SmartUPP</td>
<td>PU prevention and treatment</td>
<td>No</td>
<td>No</td>
<td>Viacore IT</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Free</td>
<td>Spain</td>
</tr>
<tr>
<td>Trata la UPP</td>
<td>Provides information for PU treatment</td>
<td>No</td>
<td>No</td>
<td>Head Life APP</td>
<td>Android</td>
<td>Tablet and smartphone</td>
<td>Spanish</td>
<td>Free</td>
<td>INA</td>
</tr>
<tr>
<td>VAPUR – Pressure Ulcer Resource</td>
<td>PU prevention</td>
<td>No</td>
<td>No</td>
<td>US Department of Veterans Affairs</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Free</td>
<td>USA</td>
</tr>
<tr>
<td>WoundMAP PUMP</td>
<td>Evaluation of PUs and development of a care plan</td>
<td>Yes</td>
<td>No</td>
<td>Mobile Health Ware</td>
<td>iOS</td>
<td>INA love for at INA</td>
<td>English</td>
<td>INA</td>
<td>INA</td>
</tr>
<tr>
<td>Ulcercare</td>
<td>PU risk assessment and recommendations of care. Patient data can be shared over the internet</td>
<td>Yes</td>
<td>No</td>
<td>Dermtap</td>
<td>iOS</td>
<td>INA love for at INA</td>
<td>English</td>
<td>INA</td>
<td>INA</td>
</tr>
<tr>
<td>Wound Mender</td>
<td>PU risk assessment. Patient data can be shared over the internet.</td>
<td>Yes</td>
<td>No</td>
<td>IOSTREAM</td>
<td>iOS</td>
<td>INA love for at INA</td>
<td>English</td>
<td>INA</td>
<td>INA</td>
</tr>
<tr>
<td>BCX Braden</td>
<td>PU prevention using Braden scale</td>
<td>No</td>
<td>No</td>
<td>BioCapax Technologies SLU</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English</td>
<td>Free</td>
<td>Spain</td>
</tr>
<tr>
<td>App</td>
<td>Description</td>
<td>Capture photos of the wound (yes/no)</td>
<td>Research involving the app (yes/no)</td>
<td>Developer</td>
<td>Platform</td>
<td>Device</td>
<td>Language</td>
<td>Paid or free</td>
<td>Country of origin</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>--------------------------------------</td>
<td>-------------------------------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>--------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Care of Sweden</td>
<td>PU prevention</td>
<td>Yes</td>
<td>No</td>
<td>Care of Sweden</td>
<td>iOS and Android</td>
<td>Tablet and smartphone</td>
<td>English, Swedish, Norwegian, Danish, Finnish, German, Spanish, French, and Dutch</td>
<td>INA</td>
<td>Sweden</td>
</tr>
</tbody>
</table>

aPU: pressure ulcer.
bINA: information not available.
<table>
<thead>
<tr>
<th>Author/year</th>
<th>Study design</th>
<th>Objective</th>
<th>Sample</th>
<th>Technology description</th>
<th>Main results</th>
<th>Characteristics of the quality of the software product ISO/IEC 25010:2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vos-Drapet al, 2013 [9]</td>
<td>Not described</td>
<td>Develop a prototype seat pressure mapping system through a mat that transmits data to a smartphone app in real time.</td>
<td>Five individuals presenting spinal cord injury and wheelchair users.</td>
<td>Each individual performed three separate sessions and sat on the mat for 3 consecutive hours. Pressure relief was performed for 2 min every 30 min.</td>
<td>Over 3 hours, the mean pressure tended to increase with time, while the dispersion index remained more constant. The app prototype did not allow the clinician to select individual scatter index areas on the map, so generic sections were used. The app allows users to self-monitor.</td>
<td>Functionality</td>
</tr>
<tr>
<td>Faux et al, 2016 [13]</td>
<td>Not described</td>
<td>Propose a smartphone app for wound tracking.</td>
<td>One individual simulates the development of a heel PU during a 7-week period of hospitalization.</td>
<td>A mobile app (prototype) evaluated by health care professionals in a controlled environment. Photos of wound evolution at 7 weeks were simulated using texture modeling. To preserve an unchanging angle of the camera shooting a transparent image, a mask was placed over the current capture of the photo instead of dots and lines around the wound.</td>
<td>The captured photos showed almost the same scale and orientation throughout the 7 weeks of the study. Quantitative results showed a variation of 40% of the area and 25% of the perimeter due to the difficulty of aligning the mask over the current image of the wound, especially in the intensive care unit or hospital room when patient mobility is reduced.</td>
<td>Functionality, efficiency, and usability</td>
</tr>
<tr>
<td>Tines 2015 [36]</td>
<td>Applied research</td>
<td>Develop a mobile app prototype that assists in the prevention and classification of PUs.</td>
<td>Eight nursing specialists and eight computer specialists.</td>
<td>It used Android, Java programming language, which was provided by Android SDK and Android Studio, and the KNN algorithm. The requirement analysis was used for software development.</td>
<td>App navigation flowchart. The app presents the user with a list of PU care recommendations. The user can capture a photo of the PU, and the system will process this image with a suggestion of the probable stage of the injury. Additionally, it calculates the score using Braden scale.</td>
<td>Functionality, efficiency, usability, reliability, maintainability, and portability.</td>
</tr>
<tr>
<td>Friesen et al, 2013 [3]</td>
<td>Not described</td>
<td>Create an interface that maximizes user compliance and data value for primary users.</td>
<td>Eight nurses from a health unit.</td>
<td>The nurses received a smartphone or tablet with the app, 90-min training, and a training manual. The nurses used the app in their daily practice with at least seven consecutive shifts. After 3 weeks, web-based research was applied on the design and functionality of the app and 6 weeks after a focus group session was held.</td>
<td>The nurses reported that the app was logical. However, they identified the need for more cross browsing between the various areas of the app and indicated that the list of treatments section was very long. They observed that the value of the wound image depended on how the photographs were taken. Based on user testing, the researchers will work on improvements in the design and development of image analysis algorithms.</td>
<td>Functionality, efficiency, usability, safety, and maintainability.</td>
</tr>
<tr>
<td>Poon et al, 2015 [2]</td>
<td>Not described</td>
<td>To develop an algorithm that determines the size of the wound in relative and absolute terms and to analyze the color of the PU image.</td>
<td>This is the SMARTWOUNDCCARE mobile app, with a description of the enhancement of image analysis algorithms.</td>
<td>The following three algorithms were used: mask image, camera calibration, and color analysis.</td>
<td>It was possible to automatically detect the size of the PU, as well as the color of the wound and to ultimately correlate the PU stage. However, it was not possible to determine the depth of the PU.</td>
<td>Functionality and efficiency</td>
</tr>
</tbody>
</table>
Application Quality Assessment

The study by Tibes [36] evaluated the largest number of software quality characteristics and described functionality, reliability, usability, efficiency, maintainability, and portability. The study by Faux et al [13] evaluated functionality, usability, and efficiency. The study by Friesen et al [3] evaluated the characteristics regarding usability, efficiency, safety, maintenance, and functionality. The study by Vos-Draper et al [9] evaluated functionality, and the study by Poon et al [2] evaluated functionality and efficiency. The study by Pérez-Barreno et al [31] did not describe software quality characteristics.

Discussion

Principal Findings

The objective of this study was to conduct a systematic review of published studies on mobile apps and a systematic survey in web-based stores looking for apps developed to identify, evaluate, treat, and/or prevent pressure ulcers in adults, and to evaluate the apps based on software quality characteristics. The studies were evaluated based on the eight software quality characteristics recommended by ISO/IEC 25010:2011 (functionality, compatibility, reliability, usability, efficiency, maintainability, safety, and portability) [34]. Based on observation of the six selected studies, we verified the use of similar technologies but different study designs and results in the initial stages, making definite analysis of app quality impossible.


Vos-Draper et al [9] developed a prototype seat pressure mapping system through a mat that transmits real-time data to a smartphone app. The pressure mat was tested for skin safety, and preliminary variables were investigated for reproducibility. In the study, the development steps of the web-based app were not described. The authors reported only that the app prototype did not allow the clinician to select dispersion index areas of the individual pressures on the map, so generic sections were used, thereby compromising app efficiency (ie, the amount of resources used in the software does not meet the user’s requirements) [34]. The authors concluded that the app was successfully developed and received by users and displayed wireless carpet pressure data on a personal smartphone, allowing users to self-monitor seat interface pressure outside the clinical setting, complying with the functionality characteristic, although the evaluation of this feature has not been described in the article. The authors suggest future tests to improve app settings and additional research to determine if the prototype can successfully modify users’ behavior in pressure relief, and there is intention to use the data provided in an individualized way. It is observed that this study is still under development. Although functionality can be considered positive by wheelchair users, it is necessary to submit the app for assessment of other quality characteristics, as the app has the potential to contribute to pressure ulcer prevention and the quality of life of wheelchair users.

According to Matthew-Maich et al [37], the successful design and development process of an app involves the continuous participation of end users. In this way, researchers and engineers will have an understanding of the context in which the solution will be used by a diverse group of end users. This also makes it possible to establish, from the onset, the specific software and hardware resources that can be considered acceptable, preferable, and compatible with the users’ needs and that will influence users’ adherence at the end of the process.

Faux et al [13] described a user-centered app design with the practical goals of usability and efficiency. The mobile app...
The app prototype was evaluated by health professionals (nurses were cited in the context of image capture) in a laboratory-controlled environment replicating in detail a hospital room or home. The experiment simulated the evolution of a pressure ulcer on the heel of a fictitious patient hospitalized in the hospital for 7 weeks. The images of wounds simulated using texture modeling were used to evaluate the app. A mask image was used in the capture of the images, and algorithms were adopted to evaluate the evolution phases of the wounds. The quantitative results indicated a variation of about 40% for the area and 25% for the perimeter of the wound image due to difficulty in aligning the mask image on the current image. The results of the laboratory experience show that evaluation characteristics, such as usability, efficiency, and functionality, are clearly insufficient in this context, and it is necessary to improve the app and perform future clinical validation with health professionals in a real-life setting with real patients. It is important for apps to be tested in real environments, because in those conditions, tests reveal information not listed during app development [38]. Researchers suggest conducting targeted surveys to assess the quality of apps in partnership with patients, health care providers, and the digital industry, according to a well-established and rigorous scientific methodology with consideration of the steps developed in the elaboration of apps, as well as their validation [38-46].

The study by Tibes [36] described the development of a mobile app prototype (UpCare) aimed to provide personalized information about each patient regarding the risk, prevention, and/or classification of pressure ulcers. The development of the app was divided into five stages, namely requirements analysis, knowledge definition, computational representation, system coding, and system evaluation. In this study, eight pressure ulcer images of only one anatomical region were initially taken from the National Pressure Ulcer Advisory Panel site for the construction of the database. The RGB (red, green, and blue) color system was used for image processing, and the K-nearest neighbor algorithm was used for classification. On providing a new pressure ulcer image to the system, the algorithm automatically identifies the most similar image in the image bank and thus can estimate the pressure ulcer stage of this image. In the end, eight nursing specialists and eight computer specialists judged the app prototype through two online questionnaires. As for the quality requirements of the app, the researcher mentioned that the following six characteristics were evaluated: functionality, reliability, usability, efficiency, maintainability, and portability. The evaluators answered the evaluation questionnaire through four case studies (two for the nursing specialists and two for the computer specialists). The case studies guided a fictitious appraisal using the app. In this way, the quality evaluations performed were compromised, as ideal evaluations should occur in real environments [38,47] and without directing results favorable to research by means of case studies. Another factor considered unfavorable is that the evaluation was performed through four case studies but the result analysis was performed jointly. For security reasons, when initializing the app, user login and password were requested; however, the security feature was not evaluated in the study. The author mentioned that the objectives for the future will be the development of the final version of the app as a product and its evaluation together with users in real practice, as well as its development for the iOS platform. Other researchers [48] note the relevance of having the app assessment performed by the final users, demonstrating that these evaluations contribute to the improvement of the app and suggestions for future work. According to Sá et al [45], when users assess an app, their interaction with the product is strengthened, often identifying needs for improvements not anticipated in the initial design.

Previous researchers [3] undertook a user test to obtain feedback on the design and functionality of the pressure ulcer monitoring app called SMARTWOUNDCARE. The app was developed for smartphones and tablets and was tested by eight nurses from a health care center. The user test encompassed a focus group and a web-based survey after a period of training and use of the app. All parameters were graded on a Likert scale with scores from 1.0 (low) to 5.0 (high). Based on the assessment, nurses reported a high degree of ease in how the app guided users to insert a new medical record (score 4.57/5.00), find an existing patient record (score 4.71/5.00), add a new wound to an existing patient record (score 4.50/5.00), evaluate a wound for the first time (score 4.57/5.00), and evaluate an existing wound that has already been evaluated (score 4.29/5.00), with strong correlations between paper forms and/or app in terms of content and expected data entry with scores of 4.60 (out of 5.00) for the Braden scale and 4.57 (out of 5.00) for the Pressure Ulcer Scale for Healing. These results show that the evaluators were able to validate quality characteristics related to functionality, usability, and efficiency. To ensure the privacy of patient information, data were stored on the device instead of a central server with remote access. However, the authors did not describe whether they were successful with respect to the app security feature. Appraisers suggested improvements in the maintenance feature, thus allowing the inclusion and exclusion of app patients, but overall, reported that the app is easy to understand and navigation is in accordance with the reality experienced in a health center. In this way, the quality characteristics of the app were evaluated positively. However, the main contribution of the app was related to the incorporation of the images of wounds (photographs) in the records of the patient, with a positive impact on caregivers’ work, health professionals, the patient, and family members. On the other hand, the researchers recognized that it is necessary to develop image analysis algorithms to detect wound size and chromaticity, thus improving the reliability characteristics of the app.

This study is the continuation of previous work, where the authors proposed to improve some of the fragilities referring to the quality of pressure ulcer images in the SMARTWOUNDCARE app [2]. The research team used machine learning algorithms and an image library to correlate the color of the wound with the stage of the wound. In order to increase user robustness, all images were processed. For the improvement of the app, three different algorithms were proposed. The first component presented is a mask image; with this element, the objective was to determine the change in the image size of the pressure ulcer on referring to a prior image of the same wound. The authors reported that the errors inherent in this method are associated with the user, including the mask’s
sharpness and the ability to align the mask over the wound during image acquisition. In this respect, the algorithm cannot achieve the objectives of the features of functionality and efficiency because the results obtained are not compatible with what is expected [34]. The second component is camera calibration. The grabcut algorithm plays an important role in this app as an image segmentation method. To estimate the size, the method compares the pixels from two images and the algorithm calculates the relative size change between these images. Compared to other algorithms, grabcut provides efficient results with minimal human interaction, and this constitutes the main benefit in this work. In this context, the second algorithm satisfactorily reaches the evaluation of the quality of efficiency [34]. The camera calibration method can be used in conjunction with the image mask method to obtain the actual size of the wound. However, the mask image and calibration component of the camera do not identify the depth and volume of the wound. Color analysis is the third algorithm, and it determines the range of colors present in an image. This can help determine the depth or stage of the pressure ulcer. According to the researchers, an inherent problem in this method depends on a certain aspect on the user’s definition of colors, and therefore, it is recommended to use a large data set to define parameters and to associate this module with machine learning. After extracting the color, the results can be sent to a specialist system to determine the pressure ulcer stage. The authors report that future work will focus on the formation of a specialist system and the development of machine learning elements, such as the support vector machine, to help determine the pressure ulcer stage. Based on the findings mentioned by the researchers, two indicators for the evaluation of quality characteristics, such as functionality and efficiency, were identified, although none of them were described in the article.

According to Pressman [49], the quality of the product depends on the quality of the development process, so it is common for higher quality apps to go through improvements in the software development process. According to Matthew-Maich et al [37], interdisciplinary app development teams need to consider specific factors when designing, deploying, and evaluating such technologies, considering working “with” and not “for” end users. With effective and efficient evidence-based app development, mHealth solutions offer great potential for improving end-user health.

The article by Pérez-Barreno et al [31] describes the GuiaUPP app. For the development of the content of the app, the researchers carried out a bibliographic search according to the main Spanish Clinical Practice Guidelines. According to the authors, the app addresses the entire problem related to the development and prevention of pressure ulcers, which can be subdivided into classification, evaluation, treatment, products, and bibliographical references. The authors point out that the GuiaUPP app provides the best and most updated evidence available on the prevention and treatment of pressure ulcers. Presently, the app is under evaluation with the intention of being included in the Mobile Health App Catalog of the Health Quality Agency of the Council of Equality, Health and Social Policies of Spain. In the app description, there is no evidence of assessment of any of the quality features of the software. The authors [38,50] argue for the need to evaluate the quality characteristics of the software, especially those of functionality, usability and efficiency. They defend that these evaluations can contribute to improvements related to appropriate technical content, better graphic presentations, and performance of the app, pointing out to specific issues and consistency with reality. According to these authors, apps should be tested in the early versions, thus optimizing the identification of problems and improving the apps before they are marketed.

**Strengths**

A positive point of this review is that it highlights the importance of a dialogue between users and researchers for the technical improvement of an app. According to Friesen et al [3], after the user test, professionals reported the importance of showing images of pressure ulcers that patients were not able to see (eg, wounds on the back of the body and under the feet), and this contributed to a better understanding by patients and family members in the fulfillment of the wound treatment guidelines. The images facilitated consultation with other health professionals and saved time by avoiding the removal and replacement of the dressing each time the doctor or specialist needed to see the wound. Based on this information, researchers [2,3] improved the algorithms of image analysis, and within this perspective, they gained new insights, so that in future phases, they may propose to incorporate artificial intelligence to assist in the classification of pressure ulcer stages. In a study carried out by Vos-Draper et al [9], important final requirements were defined for the improvement of the app prototype after discussions by a focus group.

**Limitations**

There are few available apps for pressure ulcers that have research regarding their development and evaluation. In this review, only six studies were found; however, we identified a wider range of apps (commercial names) that address pressure ulcers, with only one commercial app having research regarding its development. The studies in this review show great variability in relation to the methodology used in app development. Of the six studies evaluated, only one followed a specific method of software analysis [49]. Additionally, we need to mention the absence of a gold standard for comparison of the apps. Other limitations are related to fictitious evaluations that compromise the entire process of developing the app, as it requires evaluations with real patients and without image makeup (choosing the best image).

Another weakness is the composition of the image bank of photographs of only one anatomical region, which prevents the extrapolation of the research results to other anatomical zones. Other downsides are the lack of reporting of patient data security in apps and lack of reporting of the content sources of most apps. Pérez-Barreno et al [31] highlighted the relevance of evidence-based clinical practice, and within this context, they indicated the importance of recommendations that are developed systematically and with scientific rigor for apps in order to help professionals and patients make decisions about health care more appropriately.

http://mhealth.jmir.org/2020/6/e14266/
Clinical Implications and Future Directions

The present review identified that studies involve the initial phase of app development or improvement; therefore, apps still need to be evaluated through software quality characteristics to improve weak or absent aspects. However, the evaluation of software quality characteristics implicitly and explicitly in studies showed important points to be considered in future research. For example, capturing photographs for the identification, evaluation, and monitoring of the process of deterioration or cure of a pressure ulcer is a primordial element in app construction; however, this is an obstacle to be overcome. It is expected that the increase in technology incorporated into smartphone cameras, such as infrared thermography, algorithm enhancement, and use of artificial intelligence, may show promising results in future research. In addition, it is suggested that future studies may assess apps with regard to the technical quality of software development, using research with a rigorous scientific methodological design and with real patients. Studies should also take into account questions regarding data security, ethical issues, and the source of content, which should be based on the best scientific evidence available. It is suggested that the app image database should be composed of photographs of different anatomical regions where pressure ulcers can develop, so that the comparison of a newly captured image will be valid. In addition, there should be access to a large database of images like big data to compare newly captured images. Moreover, it is essential that apps take into account different ethnicities, with improvement of the algorithms of image analysis for individuals with darker skin, because algorithms can misclassify pressure ulcer stages in these individuals.

Conclusions

In this review, it was observed that all studies involved the initial phase of app development or improvement, and therefore, pressure ulcer apps still need to be evaluated using different software quality characteristics, so that in the future, a gold standard can be approached. Therefore, the prescription of an app for the identification, evaluation, treatment, and/or prevention of pressure ulcers in adults is currently limited. However, the evaluated studies provided important insights for future research. It is of utmost importance that future surveys develop apps jointly with users through collaborative and cocreative processes and assess patients in real-world situations across different service settings, and they should consider different ethnicities so that apps are useful to end users, such as patients, family members, health professionals, and students, in the health area. In addition, it is necessary for studies to describe the methodological course of app development in a clear and objective way in order to ensure reproducibility of the study and to offer inputs to allow future research to approach the development of ideal apps that are geared to positively impact the health of end users.

Acknowledgments

This study was funded in part by the Coordination of Improvement of Higher Level Personnel-Brazil (CAPES) (Finance Code 001) and by the National Council of Scientific and Technological Development-Brazil (CNPq) (Doctorate GD).

Authors’ Contributions

JK and MVB are the main authors of this review, and they participated in all stages of the review. JK, MVB, BEPC, and LMK wrote the protocol and commented on the draft review. MVB and JK selected the studies, conducted the quality assessment of eligible studies, and extracted data from the included studies. PRHM and CB carried out the search for studies in the references of the included studies in the review, the search of the gray literature, and the search of websites and stores of online apps. JK, MVB, ATFK, LMLD, MVMP, LGP, MC, RF, and VDT evaluated the included studies regarding software quality. MVB and JK performed the data synthesis and wrote the review draft.

Conflicts of Interest

None declared.

References


36. Tibes CM. Dissertação de Mestrado apresentada a Universidade Federal de São Carlos, São Paulo. 2015 Dec 01. Aplicativo móvel para prevenção e classificação de úlceras por pressão URL: https://repositorio.ufscar.br/bitstream/handle/ufscar/3287/6796.pdf?sequence=1&isAllowed=y [accessed 2020-06-06] [WebCite Cache ID 77DLaog6r]


38. Domingos Mello GR, Lorenzini Erdmann A, Lima Pestana Magalhães A. SEPSISCARE: AVALIAÇÃO DE APLICATIVO MÓVEL NO CUIDADO DE ENFERMAGEM AO PACIENTE COM SEPS. Cogitate Enferm 2018 Apr 26;23(2) [FREE Full text] [doi: 10.5380/cc.v23i1.52283]


Abbreviations

ICU: intensive care unit
IEC: International Electrotechnical Commission
ISO: International Organization for Standardization

©Janine Koepp, Miriam Viviane Baron, Paulo Ricardo Hernandez Martins, Cristine Brandenburg, Ariane Tieko Frare Kira, Vanessa Devens Trindade, Luis Manuel Ley Dominguez, Marcelo Carneiro, Rejane Frozza, Lia Gonçalves Possuelo, Marcus Vinicius De Mello Pinto, Liane Mahlmann Kipper, Bartira Ercilia Pinheiro da Costa. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 16.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Review and Selection of Online Resources for Carers of Frail Adults or Older People in Five European Countries: Mixed-Methods Study

Roberta Papa¹, BSc; Areti Efthymiou², PhD; Giovanni Lamura¹, PhD; Flavia Piccinini¹, MPsy; Giulia Onorati¹, MSc; Evridiki Papastavrou², PhD; Theologia Tsitsi³, MSc; Giulia Casu³⁴, PhD; Licia Boccaletti³, MSc; Alessandra Manattini¹, MSc; Rita Seneca³, MSc; Carlos Vaz de Carvalho⁵, PhD; Rita Durão⁵, MSc; Francesco Barbabella¹⁶, PhD; Frida Andréasson⁶, MA (Soc); Lennart Magnusson⁶,⁷, PhD; Elizabeth Hanson⁶,⁷, PhD

¹Centre for Socio-Economic Research on Ageing, IRCCS INRCA, Ancona, Italy
²Department of Nursing, Faculty of Health Sciences, Cyprus University of Technology, Limassol, Cyprus
³Department of Psychology, University of Bologna, Bologna, Italy
⁴Anziani e Non Solo – social cooperative, Carpi, Italy
⁵Virtual Campus, Lda, Porto, Portugal
⁶Department of Health and Caring Sciences, Linnaeus University, Kalmar, Sweden
⁷Swedish Family Care Competence Centre, Kalmar, Sweden

Corresponding Author:
Roberta Papa, BSc
Centre for Socio-Economic Research on Ageing
IRCCS INRCA
Via Santa Margherita 5
Ancona, 60124
Italy
Phone: 39 +390718004602
Email: r.papa@inrca.it

Abstract

Background: Informal carers have a crucial role in the care of older people, but they are at risk of social isolation and psychological exhaustion. Web-based services like apps and websites are increasingly used to support informal carers in addressing some of their needs and tasks, such as health monitoring of their loved ones, information and communication, and stress management. Despite the growing number of available solutions, the lack of knowledge or skills of carers about the solutions often prevent their usage.

Objective: This study aimed to review and select apps and websites offering functionalities useful for informal carers of frail adults or older people in 5 European countries (Cyprus, Greece, Italy, Portugal, and Sweden).

Methods: A systematic online search was conducted from January 2017 to mid-March 2017 using selected keywords, followed by an assessment based on a set of commonly agreed criteria and standardized tools. Selected resources were rated and classified in terms of scope. Focus groups with informal carers were conducted to validate the list and the classification of resources. The activities were conducted in parallel in the participating countries using common protocols and guidelines, a standardization process, and scheduled group discussions.

Results: From a total of 406 eligible resources retrieved, 138 apps and 86 websites met the inclusion criteria. Half of the selected resources (109/224, 48.7%) were disease-specific, and the remaining resources included information and utilities on a variety of themes. Only 38 resources (38/224, 17.0%) were devoted specifically to carers, addressing the management of health disturbances and diseases of the care recipient and focusing primarily on neurodegenerative diseases. Focus groups with the carers showed that almost all participants had no previous knowledge of any resource specifically targeting carers, even if interest was expressed towards carer-focused resources. The main barriers for using the resources were low digital skills of the carers and reliability of health-related apps and websites. Results of the focus groups led to a new taxonomy of the resources, comprising 4 categories: carer’s wellbeing, managing health and diseases of the care recipient, useful contacts, and technologies for eldercare.
Conclusions: The review process allowed the identification of online resources of good quality. However, these resources are still scarce due to a lack of reliability and usability that prevent users from properly benefiting from most of the resources. The involvement of end users provided added value to the resource classification and highlighted the gap between the potential benefits from using information and communication technologies and the real use of online resources by carers.

(JMIR Mhealth Uhealth 2020;8(6):e14618) doi:10.2196/14618

KEYWORDS
informal carers; mobile apps; websites; usability; reliability

Introduction

Informal carers are people who provide unpaid care to someone with a chronic illness, disability, or other long-lasting health or care needs outside of a professional or formal framework [1]. They represent an inherent and indispensable component of current health and social care provision across Europe, providing 80% of all long-term care [2]. Caring can be highly rewarding, but also demanding, resulting in social isolation, physical exhaustion, and psychological exhaustion, including anxiety, depression, frustration, anger, guilt, grief, stigma, and difficulties in reconciliation of work and care responsibilities [3-5]. Providing care for over 10 years and more than 40 hours per week is not a rare phenomenon, and it affects not only the carers’ physical, emotional, social, spiritual, and financial wellbeing [6] but also the quality of the care provision itself [5]. A recent estimate of the outstanding role of informal care globally, based on the prevalence and incidence of noncommunicable diseases, has highlighted that for one care recipient, there are at least 3 carers [7]. Informal care is common throughout Europe, although with different characteristics [8].

Advances in clinical research and technological innovation in health care have opened new horizons in care provision and the support of vulnerable groups, and relevant solutions are offered both through care service providers and at the patient’s home [9]. Support services based on information and communication technologies (ICTs), such as interactive services, psychoeducational and stress management programs, carers’ platforms, e-learning courses, telemedicine, and telehealth [10-18], have the potential to support informal carers in their daily tasks [19]. Nowadays, terms such as electronic health (eHealth) and mobile health (mHealth) are gaining increased attention in the research community, making their appearance in published papers for over a decade now [20]. According to the Global Observatory for eHealth [21], mHealth is defined as “a medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices.” Mobile devices can be of great advantage for carers as they are widely available and normally easier to use than PCs. They are also user-friendly and allow handy access to internet-based applications. It is worth mentioning that, in 2008, only 50% of the world population owned a smartphone [22], while in 2016, mobile web browsing overtook PCs for the first time [23].

Currently, there are apps already available on the market that can be useful for carers at any time, in any context, and for a variety of tasks. A taxonomy recently presented by Grossman et al [24] identified 8 app categories: information and resources (eg, disease information, videos, databases); useful resources for reminding of tasks and deadlines (eg, activity monitoring, personal organizer, medication reminders, diaries); support for carers (eg, support groups and chats); safety apps (eg, GPS, alarms, reminders); communication with family, friends, and professionals (eg, sharing task calendars, social networking, email, chat); care recipients’ activities (eg, reminiscence for persons with cognitive impairment, music, recreational activities, memory aids); personal health record tracking; and problem-solving solutions (eg, managing behavioral disorders).

Even though carers could benefit from many free or low-cost apps already offered on the market, some barriers prevent their use. Carers are often not aware of the apps, do not know how to install or use them on their mobile devices, or have not realized the potential benefits they could gain by using them [25-27]. Several factors are associated with the use of health-related web-based services and interventions among carers and might include the accessibility of the internet and related equipment, carers’ personal characteristics, social network and support, carers’ beliefs, duration of care, and type of web-based use (ie, reflective or passive use) [18,28]. The presence of a vast multitude of health-related apps and websites can be challenging, as it might be difficult to find reliable resources [29,30], especially for users with lower levels of eHealth literacy or mHealth literacy [31,32]. Digital skills training programs for carers are mainly available through projects, and it is not easy to find relevant publications apart from press releases and training curricula on project websites or carer associations.

Taking these considerations into account, the “Apps4Carers” study, funded by the Erasmus+ program in 2016-2018, aimed to overcome the barriers that currently limit informal carers and their care recipients from fully benefiting from learning, accessing information, and social participation opportunities offered via mobile devices. The specific objectives of this project were to: (1) select (among those already available on the market for free or at a very low cost) online resources (ie, apps and websites) offering functionalities useful for informal carers; (2) develop training content and methodologies to empower carers to use these resources; (3) provide informal carers with ICT skills for using mobile devices and online resources; (4) develop a mobile app to be used as a compact, usable, and informative library of the selected mobile resources dedicated to carers. In order to reach a greater impact and enhance transferability of the results, these aims were pursued transnationally in 5 countries (Cyprus, Greece, Italy, Portugal, Sweden), because caring is a European-wide issue, and the presence of different European country contexts, such as...
southern European and Scandinavian contexts, allowed the project outcomes to be tested in different cultural and socioeconomic contexts. This paper focused on the first stage of the project, namely the review and selection of the mobile resources to be included in the mobile app, through a mixed-methods approach based on a review and focus groups carried out in the involved countries.

Methods

Design

We used a mixed-methods approach with a sequential design. First, we searched, reviewed, and selected available online resources based on a set of criteria, followed by quantitative data collection. Then, we “validated” this list with a qualitative study (ie, focus groups) involving the end users, such as the carers. The quantitative study aimed to identify the available resources for carers and explore which features they have, which needs they address, and how usable and reliable they are. The qualitative study was used to check if the selected resources were known by the carers, if they might be useful for them, and if their classification was clear. These two phases were linked to the development of the Apps4Carers app [33], which has been designed to be an online, easy-to-use library including a peer-reviewed set of resources organized by and covering main carers’ needs. Indeed, the definition of the search strategy and the selection criteria aimed to obtain a list of resources as comprehensive as possible and of acceptable quality.

For the purpose of this study, we considered only carers aged ≥18 years and both adults and older persons as care recipients.

Search Strategy

We used a systematic approach to search online resources. In order to find apps and websites useful for carers’ tasks, 3 eligibility criteria were defined: target group, scope, and disease/condition. Regarding the target group, the search was aimed at finding resources specifically designed for carers. However, after a preliminary explorative search carried out in participating countries yielded few results, we decided to cover missing aspects considering eligible resources designed to facilitate the self-management of a disease by the patient and those not related to a specific disease or target group but supporting relevant tasks. The resources had to address the most important areas of needs and preferences of informal carers [34], such as care planning and management, information and microlearning, and communication and social inclusion, referred to as the scope. Finally, the search was focused on the major chronic diseases/conditions of the older population [35-37] associated with carer burden [38-44], selected through a rapid review of the literature. Thus, we included the following 11 chronic conditions: cardiovascular diseases, stroke, respiratory diseases, mental illness, neurodegenerative diseases, cancer, digestive diseases, sensory organ diseases, musculoskeletal diseases, diabetes, and urination incontinence. The review and selection of resources were conducted in parallel in 5 countries (Italy, Greece, Cyprus, Portugal, Sweden) from January 2017 to mid-March 2017. We used the Google search engine to look for websites, Play Store to look for Android apps, and App Store to look for iOS apps. We defined a list of keywords (see Multimedia Appendix 1) in the English language for each eligibility criterion (ie, target group, scope, and chronic disease/condition) to be used alone or in combination (eg, target AND scope, target AND disease, scope AND disease, target AND scope AND disease). Each partner was responsible for translating the keywords into their national language.

Selection Criteria

Resources identified through the search were screened for the eligibility criteria and, if eligible, assessed using a set of inclusion/exclusion and additional evaluation criteria (Textbox 1).

The resources were selected if they were available in the national languages of the participating countries (ie, Greek, Italian, Portuguese, Swedish). Researchers from Greece, Cyprus, and Portugal also searched and selected English resources, because in these countries there was a paucity of applications for carers and a large number of care workers speak English, thus could benefit from these resources. Apps were included only if they were free or available at a very low cost, defined as a price ≤€3. Websites were selected only if they were responsive for mobile devices (ie, they were designed to be optimized and easy to read on these devices). Responsiveness was assessed using the Google test tool [45].

Resources were evaluated for their update status, excluding resources that were clearly out-of-date. This assessment considered the content of the app or website, rather than a specific timeframe: For example, information about services and laws should be updated regularly, while an app about dietary habits may not necessarily be updated regularly, yet could still remain valid.

Usability was assessed with standardized tools. Apps were evaluated using the Mobile App Rating Scale (MARS) [46], a multidimensional scale developed to rate and measure the quality of health-related apps. The scale is composed of 23 items organized in 5 subscales (4 related to objective quality criteria and 1 for subjective quality), a classification section for descriptive purposes, and an optional set of app-specific items. The overall app quality score is calculated using the objective quality section, which is composed of 19 items, measuring whether the app is engaging, not boring, if it works properly, and capable of providing support. Each item is rated with a score ranging from 1 (ie, inadequate) to 5 (ie, excellent); some items can be evaluated as “not applicable.” The total score ranges from 1 (ie, very bad) to 5 (ie, very good), and the apps were included if they had a score ≥3. The usability of websites was evaluated with the System Usability Scale (SUS) [47], a simple and reliable tool for measuring the usability of a variety of products and services. The scale is composed of 10 items, rated from 1 (ie, strongly disagree) to 5 (ie, strongly agree). Total scores range from 0 to 100, and websites were included if they reached a score ≥68 [47].

In addition to these criteria, the resources were checked in terms of reliability and appropriateness of the content, data security, and privacy. These aspects, due to the wide range of resources evaluated, were not always applicable and contributed to the
overall score definition, rather than producing an immediate decision of inclusion or exclusion.

Regarding reliability, health-related apps and websites were evaluated to verify whether they were certified, developed, or endorsed by relevant organizations in the field, whereas other types of resources were evaluated according to the relevance and utility for the target group. For example, we checked if health-related resources were endorsed or reviewed by established organizations (eg, Ministry of Health, medical associations, organizations like ‘Health on the Net’) or if they were included in repositories developed by European organizations (eg, myHealthApps). Regarding wellbeing and other resources, we verified if sufficient information about the developer was present or if the content was appropriate, for example, by checking if there were experts in the field that were involved in developing or revising the contents. Resources that were clearly not reliable (eg, no information about the developer, no clear content) were excluded. For-profit resources were excluded as not appropriate for the specific objectives of this study.

Data security and privacy were evaluated by checking if the app or website considered these issues and ensured their compliance. The resources were explored to verify the presence of elements such as appropriate terms of use; registration with the request to approve terms; explanation of how personal data are collected, treated, and stored; and request of permission to access to specific device features. In the case of apps, the informative page on the store was also scrutinized.

At the end of the evaluation, the reviewers assigned an overall score to each app and website, using a 5-star rating scale from 1 (ie, very bad) to 5 (ie, very good), taking into account the criteria and experience of using the resources, keeping in mind the target group of carers. This score was useful when multiple apps or websites had similar or same features, aims, and content, since it allowed us to identify the most suitable ones to be included in the app.

**Evaluation Process**

The evaluation process was conducted in parallel in the 5 countries, following common procedures and tools. Specific guidelines were developed, including the list of keywords; definition of criteria and how to rate them; and standardized tools for the usability evaluation (ie, MARS and SUS), their explanation and scoring method, and related references. Regarding the MARS scale, we suggested to the evaluators to watch the video tutorial prepared by the authors of the scale as a training tool. Moreover, a standard spreadsheet was developed for the data extraction for each resource, with prefilled formulae for the calculation of the usability criteria. The guidelines and the spreadsheet were tested by two independent researchers and approved by all participating countries. In order to standardize the evaluation procedure and find consensus on each criterion, a pilot was conducted prior to the start of the selection process. In each country, two independent evaluators assessed at least two apps and two websites, comparing results and annotating divergent issues. An online meeting was organized to discuss the pilot results, and additional indications were defined to harmonize the review. Any difficulties encountered were discussed with a senior project researcher, who checked the procedure. Moreover, regular online meetings and discussions led to reaching consensus on the evaluation procedure.

Resources found through the online search were reviewed based on the criteria summarized in Textbox 1. Websites were evaluated by navigating across pages, while apps were reviewed first through the description page in the app store, second during the installation procedure, and finally by using the apps. At the end of the review, a list of resources was compiled, and a first classification was made, according to the resource scope.
Textbox 1. Criteria for the selection of the online resources.

**Eligibility criteria**
- Target group: at least one group, preferably carers
  - Carers
  - Care recipients
  - General public
- Scope: at least one
  - Care planning and management
  - Information and microlearning
  - Communication and social inclusion
- Diseases/conditions: at least one
  - Cardiovascular diseases
  - Stroke
  - Respiratory diseases
  - Neurodegenerative diseases
  - Cancer
  - Musculoskeletal diseases
  - Diabetes
  - Mental illnesses
  - Digestive diseases
  - Sensory organ diseases
  - Urinary incontinence

**Inclusion/exclusion criteria**
- Language: at least one national language; for Greece, Cyprus, and Portugal, English resources were also eligible
  - Italian
  - Greek
  - Portuguese
  - Swedish
  - English
- Operating system and version (apps only): at least one
  - Android
  - iOS
- Price of apps or website subscriptions (apps only) ≤€3
- Responsive (websites only), as determined via assessment by the Google test tool [45]
- Update status: not being clearly out-of-date
  - Apps: version and date
  - Websites: date of the last update
- Usability
  - Apps: Mobile App Rating Scale (MARS) quality score ≥3
  - Websites: System Usability Scale (SUS) score ≥68
Additional evaluation criteria

- Reliability and appropriateness of the content: at least some information about reliability and appropriateness; not for-profit resources
- Health-related apps and websites: presence of endorsement by Health Ministry or other relevant organizations (eg, Health on the net) or included in European Directories (eg, myHealthApps Directory)
- Wellbeing and others: presence of sufficient information about developer or appropriateness of the content
- Data security and privacy, based on the content and purpose of the app or website
- Evaluation of security and privacy assured/considered (eg, terms of use, registration, information about data collection and storage)

Data Extraction and Analysis

A standard spreadsheet was used to extract the data for the eligible resources. We retrieved a set of information for each resource, such as general information (ie, country, name, link, operating system available, and tested), eligibility criteria (ie, target, scope, disease), and inclusion/exclusion criteria (ie, language, price, update status, responsiveness, MARS quality score, SUS score). For included resources, we collected additional data, such as short description, keywords, reliability of the content, data security and privacy, name of the developer, registration needed (yes/no), expert rating, and any other relevant notes. For the apps, we added information on the need of an internet connection to use the app (yes/no), MARS subjective score, and user ratings. Descriptive statistics were used to summarize the main characteristics of the selected resources, by type (ie, apps and websites) and country. Data are expressed as frequencies or mean (SD). Data analysis was performed with the statistical software SPSS.

Qualitative Study

After the selection of the resources, we conducted a qualitative study with informal carers through focus groups, a qualitative technique aimed to collect opinions, attitudes, and perceptions from a group of selected people [48]. Participants were asked to evaluate the appropriateness of resources and their classification. The list of resources was subsequently refined, and the process concluded with a list of resources classified by categories and subcategories [49].

Focus Group Methods

We conducted 8 focus groups, 2 for each participating country (ie, Cyprus, Italy, Portugal, Sweden), except Greece. Because Greece and Cyprus share the same language, have a similar cultural background, and similar carer profiles, the results obtained in Cyprus can be generalized to Greece. The activities were carried out in parallel in all countries between March 2017 and June 2017. Common guidelines were defined and followed by all countries. These included the recruitment checklist, focus group guide, questionnaire for the participants, guidelines on how to conduct a focus group, guidelines for the analysis, spreadsheet for the collection of questionnaire data, and matrices to analyze and summarize the transcriptions.

Participants were selected according to the following inclusion criteria: being an adult informal carer of a frail adult or an older care recipient; providing at least 4 hours per week of support, excluding financial support or companionship [50]; caring for a care recipient who needed support in activities of daily living (eg, mobility) or had a noncommunicable disease; being the primary carer of the care recipient; owning a smartphone or tablet; and having at least medium digital skills (ie, being able to use apps and surf the web on his or her own). These criteria were chosen to select carers with a high burden of care, representing the target group that could most benefit from the use of mobile resources. Moreover, we needed carers with at least medium-level digital skills so that they would have some prior knowledge and experiences with using mobile devices, which is the focus of the study, and would be able to provide their opinions on the points raised during the focus group.

The focus group guide included an introduction to present the project and objectives of the focus group, followed by 4 sections: (1) use of smartphone or tablet and apps; (2) evaluation of selected resources; (3) the Apps4carers visual prototype evaluation (including classification of the selected resources); and (4) training courses and material. The results presented here refer to the first and second sections and the topic of classification to the third one. The evaluation of the resources (section 2) was carried out by presenting to the carers the kind of resources included in each scope, explaining how these resources could support carers’ tasks, and providing at least 2 or 3 examples of apps or websites for each scope, by screenshots (ie, storyboard) or navigating the apps or websites in real-time (ie, walkthrough). To avoid an extensive and potentially burdensome discussion with participants, two focus groups were organized in each country, each one addressing different categories of resources and ensuring, in this way, that all categories were covered. Regarding the classification of the resources, we presented and explained the classification used and asked carers to evaluate it in terms of clarity of both the names of the categories and resources included in each one.

Recruitment of potential focus group participants was carried out through a checklist and via phone calls or face-to-face contact with participants, mainly through carers’ associations or health or social care services. In each country, suitable locations for the focus groups were selected, considering comfort and accessibility for participant carers. Immediately prior to the session, participants were requested to sign an informed consent form explaining the aims of the project and procedures. A short sociodemographic background questionnaire was also administered to participants, in order to gather the main sociodemographic characteristics of both the carers (ie, age, gender, marital status, education, working status, children, relation with care recipient) and the care recipients (ie, age, sex, location, disease, and care relation with care recipient).
duration of care, hours of care per week, living conditions, level of dependency, financial support or allowances, presence of a private care assistant).

Focus groups were conducted by qualified personnel (eg, social researchers, psychologists) and involved a moderator or facilitator leading the focus group and an assistant moderator or observer listening to the focus group's discussion and taking detailed notes. Each discussion, with the participants' permission, was audio-recorded and transcribed verbatim. Each focus group lasted about 2 hours.

**Focus Group Analysis**

Due to the different languages, focus groups were analyzed by each country project team separately using a common template, and the results were merged to produce a common report. The qualitative data were analyzed manually adopting the framework analysis technique [48,51], a case and theme-based approach that reduces data through summarization and synthesis using a matrix. This method helped to order data, allowing the researchers to analyze them both by case (ie, the carer’s point of view) and by theme. Data management and interpretation are sequential: it starts deductively from the aims and objectives of the study but reflects the original observation of the people (ie, it is “grounded” and “inducted”). Initially, focus group transcriptions were read multiple times for the researchers to become familiar with them. A set of a priori themes were defined based on the objectives of the study and the focus group guide, to be further integrated with additional themes, if necessary. A standard matrix was used to summarize and analyze data of each theme, where rows represented participants and columns the subthemes. Each country was asked to provide a summary of the results for each section and subthemes, including relevant quotations, if any.

**Results**

**Selection of Mobile Resources**

The search strategy retrieved 406 resources eligible for target group, disease/condition, and scope (Figure 1), of which 282 were apps and 124 were websites. Assessment of the inclusion and exclusion criteria and additional evaluation criteria led to the exclusion of 182 resources (144 apps and 38 websites) and the selection of 224 resources (ie, 138 apps and 86 websites; see Multimedia Appendix 2) [49]. The main reasons for exclusion of the apps were price, usability or technical issues, and update status, whereas websites were excluded primarily because of a lack of responsiveness, usability, and update status.

![Flowchart of the selection process](http://mhealth.jmir.org/2020/6/e14618/)
Characteristics of Selected Resources

The selected resources were classified into 3 categories, according to their scope. In the first category, “Care plan and management,” we included resources useful for carers in the organization and planning of daily and long-term care tasks, such as keeping track of therapy or symptoms of a disease and sharing information with other family member or with a health care professional. The second category, called “Information and microlearning,” collected resources providing information about the diseases (eg, symptoms, therapies) and about any public or private services that offer support to the care recipients and their carers. Finally, the category “Communication and social inclusion” compiled the resources providing support, advice, or sharing of information and experiences with other carers. Each resource was allowed to be related to more than one category.

The main characteristics of the selected resources are shown in Table 1.

The number of apps was quite homogeneous across countries, whereas websites varied from 5 in Portugal to 41 in Sweden. Among the selected apps, 35 were in the English language (of which 20 were in Greece/Cyprus and 15 were in Portugal). Only 17.0% (38/224) of the resources were devoted specifically to carers, whereas 22.3% (50/224) of all resources and 14.5% (20/138) and 35% (30/86) of apps and websites, respectively, addressed both carers and care recipients. As expected, apps were mainly related to issues of care planning and management, while websites provided mainly information and microlearning. Half of the resources (109/224, 48.7%) were disease-specific, specifically addressing one or more diseases/conditions of the care recipient. In most cases, these were represented (data not shown) by neurological diseases (eg, Alzheimer’s disease) for both apps (22/138, 15.9%) and websites (32/86, 37%); stroke (12/138, 8.7% and 12/86, 14%, respectively), followed by cardiovascular diseases (11/138, 8.0%) and diabetes (8/138, 5.8%) for apps; and musculoskeletal diseases (10/86, 12%) and cancer (8/86, 9%) for websites. The remaining resources not specifically addressing one disease/condition included information and utilities on a variety of themes, such as finding a pharmacy, explaining social services or legal aspects, reminders for therapies and medications, utilities for the older person, or emergency numbers.

The majority of the resources (153/224, 68.3%) were evaluated as reliable and with appropriate content, while we were able to verify the assurance of data security and privacy for only 77 resources (77/224, 34.4%). However, regarding this issue, it should be noted that a high number of resources only provided information, which in general does not imply any request or use of personal data from the user. As for the mode of use, 59.4% (82/138) of the apps required registration, and 50.7% (70/138) needed an internet connection. Expert ratings reached a mean score of 3.7 (SD 0.7) on a scale from 1 to 5. User ratings, purely related to the apps, were slightly higher, at a mean score of 4.0 (SD 0.6).
Table 1. Characteristics of the selected resources.

<table>
<thead>
<tr>
<th>Country</th>
<th>Apps (n=138), n (%)</th>
<th>Websites (n=86), n (%)</th>
<th>Total (n=224), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greece and Cyprus(^a)</td>
<td>35 (25.4)</td>
<td>16 (18.6)</td>
<td>51 (22.8)</td>
</tr>
<tr>
<td>Italy</td>
<td>42 (30.4)</td>
<td>24 (27.9)</td>
<td>66 (29.5)</td>
</tr>
<tr>
<td>Portugal(^b)</td>
<td>33 (23.9)</td>
<td>5 (5.8)</td>
<td>38 (17.0)</td>
</tr>
<tr>
<td>Sweden</td>
<td>28 (20.3)</td>
<td>41 (47.7)</td>
<td>69 (30.8)</td>
</tr>
<tr>
<td>Operating system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Android</td>
<td>22 (15.9)</td>
<td>N/A(^c)</td>
<td>N/A</td>
</tr>
<tr>
<td>iOS</td>
<td>25 (18.1)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Both</td>
<td>91 (65.9)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Target group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carers</td>
<td>25 (18.1)</td>
<td>13 (15.1)</td>
<td>38 (17.0)</td>
</tr>
<tr>
<td>Carers and care recipients</td>
<td>20 (14.5)</td>
<td>30 (34.9)</td>
<td>50 (22.3)</td>
</tr>
<tr>
<td>Care recipients</td>
<td>17 (12.3)</td>
<td>13 (15.1)</td>
<td>30 (13.4)</td>
</tr>
<tr>
<td>General public</td>
<td>76 (55.1)</td>
<td>30 (34.9)</td>
<td>106 (47.3)</td>
</tr>
<tr>
<td>Scope(^d)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care plan and management</td>
<td>72 (52.2)</td>
<td>20 (23.3)</td>
<td>92 (41.1)</td>
</tr>
<tr>
<td>Information and microlearning</td>
<td>80 (58.0)</td>
<td>73 (84.9)</td>
<td>153 (68.3)</td>
</tr>
<tr>
<td>Communication and social inclusion</td>
<td>29 (21.0)</td>
<td>31 (36.0)</td>
<td>60 (26.8)</td>
</tr>
<tr>
<td>Disease-specific (yes)</td>
<td>58 (42.0)</td>
<td>51 (59.3)</td>
<td>109 (48.7)</td>
</tr>
<tr>
<td>Reliability/appropriateness of the contents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not reliable</td>
<td>1 (0.7)</td>
<td>0 (0.0)</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>Reliable</td>
<td>91 (65.9)</td>
<td>62 (72.1)</td>
<td>153 (68.3)</td>
</tr>
<tr>
<td>Not assessed</td>
<td>46 (33.3)</td>
<td>24 (27.9)</td>
<td>70 (31.3)</td>
</tr>
<tr>
<td>Data security and privacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not assured</td>
<td>9 (6.5)</td>
<td>5 (5.8)</td>
<td>14 (6.3)</td>
</tr>
<tr>
<td>Assured</td>
<td>50 (36.2)</td>
<td>27 (31.4)</td>
<td>77 (34.4)</td>
</tr>
<tr>
<td>Not assessed or not relevant</td>
<td>79 (57.2)</td>
<td>54 (62.8)</td>
<td>133 (59.4)</td>
</tr>
<tr>
<td>Registration needed (yes)</td>
<td>82 (59.4)</td>
<td>10 (11.6)</td>
<td>92 (41.1)</td>
</tr>
<tr>
<td>Connection needed (yes)</td>
<td>70 (50.7)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Usability(^e,f)</td>
<td>138, 3.6 (0.4)</td>
<td>81, 81.5 (8.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>User rating (1-5)(^f)</td>
<td>117, 4 (0.6)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Expert rating (1-5)(^f)</td>
<td>138, 3.7 (0.7)</td>
<td>86, 3.8 (0.7)</td>
<td>224, 3.7 (0.7)</td>
</tr>
</tbody>
</table>

\(^a\)20 apps were in the English language.
\(^b\)15 apps were in the English language.
\(^c\)N/A: not applicable.
\(^d\)Each resource could have multiple scopes.
\(^e\)Usability was assessed using the Mobile App Rating Scale for the apps and the System Usability Scale for the websites.
\(^f\)Reported as n, mean (SD).

Focus Group Results

Focus groups involved 47 carers in 4 countries (ie, Cyprus, Italy, Portugal, Sweden) and aimed to evaluate the selected resources and their classification. Overall, enrolled participants had a mean age of 55 years (SD 12.0) years, were mainly female (32/47, 68%), were married (32/47, 68%), had a medium-high level of education (39/47, 83%), were employed (26/47, 55%).
and had at least one child (33/47, 70%). They were providing care primarily to a parent (29/47, 62%) or a spouse/partner (12/47, 26%). About half of the carers (19/42, 45%) provided assistance for 2–4 years, while 36% (15/42) provided assistance for longer than 5 years. More than 20 hours per week of care was provided by 48% (20/42) of participants; among these participants, 10 people were providing full-time care (10/42, 23.8%). Regarding the living situation, 17 carers (17/46, 37.0%) lived with the care recipient, while 18 (18/46, 39%) lived within walking distance. Care recipients had a mean age of 77 years (SD 9 years) and were, to a large extent, slightly or moderately dependent (32/46, 69%), while 28% (13/46) were severely dependent. Only 38% (17/45) received financial support or an allowance from the state or public organizations, and 34% (15/44) were paying for a private care assistant who provided part-time or full-time support.

The focus group results are presented in Table 2, according to themes, subthemes, and country. Three a priori themes are reported, including the use of devices and resources, evaluation of resources, and classification of resources.

Regarding the use of mobile devices (ie, smartphones, tablet), all participants owned a smartphone or a tablet, but the majority reported limited knowledge of smartphone features and functions, stating that they used their devices only for basic activities. Only the Swedish group included some expert users who were quite confident in using apps, for both daily activities and caring duties (eg, searching for health information, setting reminders, using locators). Italian and Portuguese carers reported usually asking for the help of their relatives to download, install, and use apps. In relation to resources (ie, apps, websites) dedicated to carers, almost all of the participants had no previous knowledge of resources specifically targeting carers. Most used apps and websites to support daily tasks (eg, communication, navigation, banking, social networking, games) or to find the information they needed (eg, travels, news, online shopping).

The second theme was the evaluation of the resources selected in each country, in terms of previous knowledge, interest, potential benefits, barriers, and perspective about resources in the English language. Overall, participants expressed a major interest in the resources presented during the focus groups, considering them useful tools for care provision and assistance. Some participants took note of the names or installed them immediately after the end of the discussion. Some remarks were raised by several participants, who reported that some selected resources were not always relevant for experienced carers, yet they could be deemed useful for novice carers. Another issue raised was that the resources could be useful but they do not solve practical issues of service availabilities, do not replace the need for direct or face-to-face contact with health and social care professionals, or they lack a “local” perspective.

Most of the participants did not know any of the apps or websites that were presented; however, they all seemed highly interested in their functions and potential benefits. The perceptions of the carers about the examples of resources presented were generally positive. They reported that the apps and websites seemed easy to use and intuitive. Moreover, the carers appreciated the resources for the following aspects: the relevant information provided (eg, disease explanation, tips and guidelines for specific actions or events), presence of multiple functions (eg, drug reminder and therapy management instead of a simple alarm clock), interactive approach (eg, guided exercises, personalized plans), entertainment aspect (eg, digital photo album), and possibility to share information with health and social care professionals or other informal carers. The carers reported that the availability of these resources could make them feel safe because, in some cases, these resources could be lifesaving (eg, emergency events, locators for people affected by dementia), could reduce anxiety about tasks to be performed (eg, checking drug therapy), and could facilitate care management (eg, recipes for specific conditions). An additional aspect raised by the Italian carers was the potential benefit of the use of these resources by paid carers, in order to improve the quality of care provided.

When asked about obstacles and barriers that could prevent carers from using these resources, participants indicated as their main concern the low level of digital skills of adult carers, in particular among older carers. For example, some apps were evaluated as being too complex, requiring more competency to be used, while others required the care recipients to have and use a smartphone, which could be problematic due to their low digital skills or even their fear of using new technologies. Some participants reported that the resources designed for care recipients would be more suitable for carers instead. When talking about health-related apps, some Swedish and Cypriot users raised the issue of reliability, explaining that what prevented them from using online resources is the uncertainty of the accuracy and reliability of the information provided in the apps.

With the exception of the Greek and Cypriot users, participants from other countries were not interested in apps available in languages other than their national language (ie, English language). However, multilingual resources were considered to be potentially helpful for privately paid care assistants, in most cases with a migrant background, who may have limited knowledge of the language of their host country.

As for the classification of resources, participants had some difficulties in understanding which kind of resources were included in the 3 categories (ie, care plan and management, information and microlearning, communication and social inclusion). The main concerns raised were related to the words selected to represent the categories (in each national language), which were evaluated as too general in some cases or too difficult to understand what they consisted of in other cases. The moderators elicited the provision of suggestions to improve the classification and labels associated with each category, to inform the subsequent development of the project’s app. Main suggestions referred to the use of simple and widely understandable words, possibility to receive a brief explanation of what is included in each category, and use of graphic elements.
Table 2. Focus group results.

<table>
<thead>
<tr>
<th>Theme and subthemes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of devices and resources</strong></td>
<td></td>
</tr>
<tr>
<td>Use of smartphone or tablet</td>
<td>All owned smartphones or tablets (IT(^a), CY(^b), SE(^c)), some also computers (SE); most owned smartphones/tablets, some only personal computers (PT(^d))</td>
</tr>
<tr>
<td>Digital skills</td>
<td>Half needed support to download and install apps (IT, SE); only some had difficulty with finding and downloading the apps (CY); only half were familiar with the “app concept” (PT)</td>
</tr>
<tr>
<td>App use and opinions</td>
<td>Daily tasks, communication, entertainment, and searching for information (IT, CY); basic functions (PT); communication, entertainment, paying bills, reading news; and searching for health information on the web (SE)</td>
</tr>
<tr>
<td><strong>Evaluation of resources</strong></td>
<td></td>
</tr>
<tr>
<td>Previous knowledge of the resources</td>
<td>None knew of the presented apps/websites (IT, PT); most did not know the presented apps/websites, while some were familiar with similar resources (CY); some participants had heard about some app or even had it on their smartphones, while some did not know of any resources (SE)</td>
</tr>
<tr>
<td>Interest</td>
<td>Great interest towards the resources (IT, CY, PT); generally interested, while experienced carers found the selected resources not always relevant for them (SE)</td>
</tr>
<tr>
<td>Benefits</td>
<td>Considered helpful for care provision and assistance (IT, PT, CY) and for obtaining reliable information (CY); not many were mentioned, while participants felt their caring situation was too complex for an app to be helpful (SE)</td>
</tr>
<tr>
<td>Barriers</td>
<td>Low digital skills of carers (IT, PT) and care recipients (IT); usability and reliability issues, poor Greek translations (CY); problems concerning the trustworthiness of apps (SE)</td>
</tr>
<tr>
<td>English language</td>
<td>They would not use any app in English (IT, PT, SE); English resources could be useful for immigrants and migrant paid assistants (CY)</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>Categories and labels were not clear or easy to understand (IT, CY, PT, SE)</td>
</tr>
</tbody>
</table>

\(^a\)IT: Italy.  
\(^b\)CY: Cyprus.  
\(^c\)SE: Sweden.  
\(^d\)PT: Portugal.

Final Classification of the Resources

The focus group results were used to “validate” the list of selected resources, confirming their usefulness for the carers and the need to improve awareness about the resources among carers and their organizations. However, the draft classification devised using the scope of the resources was not widely understood by the carers. Although the focus group participants were not able to provide specific suggestions for improvements, they provided valuable information about their needs and their perceptions of the resources. These elements, together with an in-depth analysis of the resources selected, supported the researchers in the definition of a new, more accessible, and coherent classification. First, we decided to highlight, by creating a separate category, the resources specifically targeted to carer wellbeing (eg, relaxation techniques). The second step was to identify, for each resource, the main feature, functionality, or aim and link it to a major carer need, to facilitate the carers, especially novices, to find the appropriate resources. For this purpose, the analysis of the resources led to the selection of three needs: management of the health status of the care recipient, connection with individuals or organizations able to support the carers in their tasks, and the need to have practical tools to overcome barriers and problems while caring (eg, communication). These considerations led to the final classification presented in Table 3, including 4 categories and, for each, a classification in subcategories, with examples provided for descriptive purposes. The first category was named “Carer wellbeing” and included resources supporting carers in dealing with stress management, such as relaxation techniques and mindfulness. The second one, called “Managing health and diseases of the care recipient,” comprised not only the resources related to care planning but also those providing information about the diseases/conditions of the care recipient. Services, helplines, and peer support were grouped under the label “Useful contacts,” whereas utilities and tools potentially helpful in the care of older people have been placed under the category “Technologies for eldercare.” In this classification, each resource was allocated to only one category. It was also decided that the online library would include a fifth category comprising all the resources with the possibility for the carers to search among them by using filters and keywords. This option aimed to make the app more accessible and suitable also for skilled carers.
Table 3. Final taxonomy of the selected resources.

<table>
<thead>
<tr>
<th>Category and subcategories</th>
<th>Kind of resources included (examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carer wellbeing</td>
<td>Tools devoted to carers (eg, relaxation, stress management)</td>
</tr>
<tr>
<td>Managing health and diseases of the care recipient</td>
<td>Information, advice and tips to manage symptoms and problems, rehabilitation tools (eg, cognitive training), assessment and tracking of parameters, sharing information among carers and professionals, prevention</td>
</tr>
<tr>
<td>Diseases and health-related issues</td>
<td>Information, reminders and management</td>
</tr>
<tr>
<td>Medications</td>
<td>Information and advice, tracking systems</td>
</tr>
<tr>
<td>Nutrition and diet</td>
<td></td>
</tr>
<tr>
<td>Useful contacts</td>
<td>Services: information, research, and location (eg, pharmacies); associations: information and events; legal and financial information and advice</td>
</tr>
<tr>
<td>Services</td>
<td>Helplines; emergencies: direct link, calls; emergencies: information; peer support and expert advice (eg, Facebook groups and fora)</td>
</tr>
<tr>
<td>Help and information</td>
<td>Resources for disabled people (eg, interfaces, assistive devices)</td>
</tr>
<tr>
<td>Technologies for elder care</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

This study aimed to review and select available mobile resources supporting informal carers of frail adults or older people in 5 European countries. The combined use of standardized tools, real tests, and focus groups provided a comprehensive review of the actual “state-of-the-art” status of mobile resources for carers. We found few resources dedicated to carers, and we covered the missing aspects with alternatives targeting care recipients or the general population. We also found that there are some underdeveloped areas and some specific features requiring attention and additional effort by the developers of apps and websites. Finally, the focus groups confirmed the usefulness of the selected resources, underlined the main barriers to their use, and led to the definition of a new taxonomy.

In greater detail, the main result emerging from this study concerned the paucity of apps and websites specifically developed for informal carers. We identified only 25 apps and 13 websites addressing this target group, aiming to provide guides to stress management; information, advice, and tips to provide care for people with different diseases (eg, dementia, Parkinson’s disease, stroke); or events and services for carers. These results are in line with a recent review conducted in the United States [24], which found only 44 apps for carers of older adults. The lack of adequate resources able to support carers is an indirect sign of the largely unrecognized role they play in everyday care for older people in our aging societies. Informal carers do provide a vast amount of caring activities and thus enhance the sustainability of welfare systems [52,53], in a context where the need for long-term care is increasing [54] and the number of people potentially available to provide care is declining [55]. At the same time, there is evidence that informal carers are at a higher risk of developing carer burden and psychiatric morbidity [56], meaning that they also need support to deal with their wellbeing and caring situation. One possible solution to this problem is to provide valuable ICT-based support tools to carers, so they can provide good quality care and receive support for themselves [19]. The potential usefulness and beneficial impact of these tools were confirmed by the carers involved in the focus groups. In the majority of cases, they were not aware of these resources but, once presented, they understood their potential and were willing to use them for their caring activities.

The needs and preferences of carers vary considerably and differ by personal characteristics of both the carer and the care recipient, the kind of relationship between them, and the stage of the caring cycle [28,57-59]. For example, when assuming this role, carers have to acquire all useful information and training related to the disease/condition affecting the care recipient. Then, they often search for communication and support services, both formal and informal. Finally, they also need support from peers or care professionals to manage the stress and burden caused by caring activities, often accumulating over time. The carers' opinions collected during the focus groups confirmed these differential needs and the necessity to identify relevant online resources able to address them. In our review, we found apps and websites covering all these aspects, even if some areas were more developed compared to others. The majority of the selected resources provided information, followed by different apps for monitoring disease(s) and medicines. In the area of peer support, we found a few, but interesting, resources (eg, Facebook groups), which are likely to increase in the future thanks to the diffusion of social networks. These resources represent a source of information exchange and support for carers and a way to increase the effectiveness of their role and reduce the stress associated with the caring activities [60].

We found a paucity of resources aiming specifically to manage stress and burden, such as guided exercises of relaxation techniques or meditation. This is a promising sector, and a recent review [61] underlined the positive effects of online mindfulness-based interventions on mental health.

Although the area of information revealed the highest number of resources, there is a need to further address issues of reliability of the content [29,30] and usability [26,62]. In many cases, the absence of indications of endorsement or revision by experts or by devoted bodies hampered even the evaluator's
assessment of the reliability of the content for both apps and websites. The focus group findings confirmed that these issues often prevented carers from using the resources. Evaluation and trust of web-based health information by users is related to eHealth literacy and mHealth literacy and is influenced by carers’ characteristics such as socioeconomic position, education, and employment [31,32,63]. Therefore, there is a need to train carers on how to evaluate the quality and accuracy of online health information and to increase the use of certified web-based resources that are easily recognizable. Second, a consistent number of resources were excluded because of poor usability and technical problems. Although we evaluated only free or very cheap resources, in many cases, paid apps were just “premium” versions of free apps, where more content or functionality are provided to the user with the same design. Moreover, when considering websites, many were excluded due to the lack of a “mobile” version. Importance of usability was also confirmed by the users, who reported the risk of not using or ceasing to use apps if they are too complex. This was the case for the classification used for the selected resources, which failed to be understood by the carers and required a complete revision.

Finally, the focus group results highlighted that carers having different characteristics and from different European contexts shared the same opinions and perspectives on the use and usefulness of mobile resources. This provides evidence of the potential impact and transferability of this initiative in other European countries. Despite the original findings presented so far, this study has some limitations, too. First, although the review and selection processes used a systematic approach, the search could have excluded some relevant resources, due to the functionality of the tools used for the search or due to the translation of the keywords in the national languages. Second, we restricted our search to the major diseases/conditions associated with a high burden of care. Although this could have missed some resources, we do consider that we covered the majority of the needs of the target group considered. Third, the evaluation of the resources in terms of update status, reliability/appropriateness of the content, and data security and privacy was carried out by the researchers following a set of general indications, thus not assuring an objective rating. Finally, given the relatively short timeframe of our project, it was not feasible for the selection of the resources to also undergo a review by external experts in all countries, which might have ensured additional validation of the obtained results.

Notwithstanding these limitations, this study allows us to state that mobile resources are potentially valuable tools for carers, as they can be used anytime and anywhere to search for information, contact persons, and learn something new and relevant about the caregiving role [64]. Moreover, these tools could also have a positive impact in terms of sustainability of the welfare system and overall cost savings [59]. This study confirmed that, although mobile resources such as apps and websites could support carers in their tasks, more effort should be spent in creating a greater awareness about these resources and in developing good-quality resources specifically targeted to carers, by addressing their needs and preferences [18]. To reach these objectives, apps and websites need to be developed and co-designed together with the end users while enhancing the acquisition of digital skills needed to use these tools. We further exploited the results of this study in the subsequent phases of the project, by developing an online library of the selected resources [33] and training materials to support carers wishing to learn how to use these tools. The main future challenge, for both research and market fields, will remain the understanding of how to best fill the gap between the potential benefits for carers of using ICTs and their real use of these resources.

Acknowledgments

This study was co-funded by the Erasmus+ programme of the European Union, under the Project “Apps for carers”, Grant Agreement n. 2016-1-SE01-KA204-022067. This study was partially supported by Ricerca Corrente funding from the Italian Ministry of Health to IRCCS INRCA.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of keywords used for searching the resources.
[DOCX File . 21 KB - mhealth_v8i6e14618_app1.docx ]

Multimedia Appendix 2
List of selected resources.
[PDF File (Adobe PDF File). 1100 KB - mhealth_v8i6e14618_app2.pdf ]

References


32. Ahmed S. IDEALS.: University of Illinois; 2017 Mar 20. mHealth literacy: characterizing people’s ability to use smartphone-based health-related applications URL: https://www.ideals.illinois.edu/handle/2142/97267 [accessed 2019-04-22] [WebCite Cache ID 77pOZzntf]


---

**Abbreviations**

- **eHealth**: electronic health.
- **ICT**: information and communication technology.
MARS: Mobile App Rating Scale.

mHealth: mobile health.

SUS: System Usability Scale.
Elderly Medication Adherence Intervention Using the My Interventional Drug-Eluting Stent Educational App: Multisite Randomized Feasibility Trial

Andrew Dallas Boyd¹, MD; Chioma Iheanyi Ndukwe¹, BSc; Anandu Dileep¹, BSc; Olivia Frances Everin¹, BSc; Yingwei Yao², PhD; Betty Welland³; Jerry Field³; Matt Baumann³, PhD; Jose D Flores Jr³; Adhir Shroff¹, MD, MSc; Vicki Groo¹, PharmD; Carolyn Dickens¹, PhD, RN; Rami Doukky⁵, MD; Regeena Francis⁵, RN; Geraldine Peacock⁵, RN; Diana J Wilkie², PhD, RN

¹Department of Biomedical and Health Information Science, University of Illinois at Chicago, Chicago, United States
²Biobehavioral Nursing Science, University of Florida, Gainesville, FL, United States
³Patient Advisor, Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, IL, United States
⁴Department of Pharmacy Practice, University of Illinois at Chicago, Chicago, IL, United States
⁵Division of Cardiology, Cook County Health, Chicago, IL, United States

Corresponding Author:
Andrew Dallas Boyd, MD
Department of Biomedical and Health Information Science
University of Illinois at Chicago
1919 W Taylor St
M/C 530
Chicago, 60612
United States
Phone: 1 3129968338
Email: boyda@uic.edu

Abstract

Background: A lifesaving treatment for myocardial infarction is the placement of a stent in a closed or obstructed coronary artery. The largest modifiable risk factor after receiving a stent is medication adherence to Dual AntiPlatelet Therapy, a combination of P2Y12 inhibitors and aspirin.

Objective: This study aimed to determine the acceptability of a protocol and an intervention using the My Interventional Drug-Eluting Stent Educational App (MyIDEA) and to evaluate medication adherence using the proportion of days covered (PDC) and platelet activation tests in a multisite randomized controlled trial.

Methods: Potential participants who received a post percutaneous coronary intervention (PCI) procedure with a drug-eluting stent were approached. All patients older than 50 years and who spoke English were recruited. Participants were recruited, baseline demographics were collected, and the Hospital Anxiety and Depression Scale (HADS), Rapid Estimate of Adult Literacy in Medicine-Short Form, Burden-Benefit questionnaire, 36-Item Short Form Health Survey, and PCI knowledge questionnaire were administered. Block randomization was used to randomize participants to either usual care or MyIDEA supplementation. MyIDEA is a personalized educational intervention based on the Kolb experiential learning theory using patient narratives for education. During the visits, participants’ blood was collected to measure platelet suppression from medication. During the second and third encounters, the Morisky medication adherence score and cardiology outcomes were measured. The study was conducted at the University of Illinois Hospital and John H Stroger Jr Cook County Hospital with appropriate ethical approvals. Platelet suppression was measured through aspirin reactive units and P2Y12 reactive units. Medication adherence was measured using the PDC. The analysis team was blinded to the participants’ group membership. The primary outcome was a feasibility analysis of recruitment and retention.

Results: The mean age of participants was 60.4 years (SD 7.1); the majority of patients were black and non-Hispanic. The majority of patients’ reading levels were seventh grade or above, and they were not very familiar with other electronic devices for information and communication. The number of control subjects was 21, and the number of participants in the interventional arm was 24. The interventional group was able to use MyIDEA in both the hospital and outpatient setting. However, there was
no significant difference in platelet suppression or medication adherence between groups. There were also differences between the groups in terms of depression and anxiety, initially, as measured by HADS. No documented adverse event associated with the intervention was found.

Conclusions: Elderly patients are willing to use tablet devices to be educated about health conditions. Additional studies are required to measure the effectiveness and determine the most suitable timing and location for patient education.

Trial Registration: ClinicalTrials.gov NCT04439864; https://clinicaltrials.gov/ct2/show/NCT04439864

(JMIR Mhealth Uhealth 2020;8(6):e15900) doi:10.2196/15900

KEYWORDS
mobile application; medication adherence; drug eluting stents; percutaneous coronary intervention; patient education; Kolb learning theory

Introduction

Background

A lifesaving treatment for myocardial infarction is the placement of a stent in a closed or an obstructed coronary artery [1]. The placement of a stent restores blood circulation to areas of the heart for patients with coronary artery disease. The most commonly used stent in the United States is the drug-eluting stent (DES) [2]. In addition, the largest modifiable risk factor after receiving a stent is medication adherence to dual antiplatelet therapy (DAPT), a combination of P2Y12 inhibitors and aspirin [3]. By stopping the medication, the risk of death is increased seven times compared with patients who follow medication directions [4-7]. The range of medication nonadherence has been reported from 7% to 50% [8,9]. Even missing only 48 hours of medication has been reported to increase the risk of stent thrombosis [10]. Current literature has shown that 5.4% of patients do not pick up their medication from the pharmacy [3]. The most common reasons for stopping their medication were a lack of understanding and not knowing how long they would need to take the medication [11-13].

A review of a number of different methods have been attempted to increase medication adherence in patients poststent [14]. A pharmacy-led intervention [15] has not been shown to increase medication adherence. Another challenge is that a lower socioeconomic status has been affiliated with lower health-seeking behavior [16]. One large area of innovation is in engaging patients in mobile health. Although currently there are over 100,000 apps to download from the Google Play and Apple App stores, very few have been studied in randomized controlled trials [17], and the apps are written at a grade level higher than the reading level of the patients [18]. The strength of using mobile health technology lies in its ability to engage the patient via customized interventions [19,20].

The My Interventional DES Educational App (MyIDEA) was a computing tablet-based education program developed with patients, cardiologists, pharmacists, nurses, informaticians, and biomedical illustrators [21]. MyIDEA was developed using the Kolb experiential learning theory such that it was customized to the patient [22]. MyIDEA also used patient narratives to help the patient plan and overcome the common barriers to medication adherence.

Objectives

This study aimed to determine the acceptability and the effect of MyIDEA medication adherence through proportion of days covered (PDC) and platelet activation tests in a multisite randomized controlled trial. MyIDEA offers an immediate response to common questions after stent procedures when the patient is at home and reviews medical instructions from nursing staff.

Methods

A feasibility study was conducted at 2 urban hospitals, the University of Illinois Hospital and John H Stroger Jr Cook County Hospital, to evaluate the medication adherence of patients availing DAPT. Both hospitals care for patients with a low socioeconomic status and have a high percentage of minority patients. The participants were recruited from University of Illinois Hospital from May 2014 to February 2016 and from John H Stroger Jr Cook County Hospital from June 2016 to February 2017.

Recruitment

The study was approved by the institutional review boards of both institutions. The enrollment criteria included patients who (1) received a DES in a percutaneous coronary intervention (PCI) during a hospitalization for the PCI procedure, (2) were older than 50 years (funding agency requirement), and (3) spoke and understood English. The exclusion criteria were as follows: (1) inability to give informed consent and (2) allergy to aspirin. The change to the study design was to include a second site. The participants were approached in person for recruitment between 4 and 24 hours after their PCI procedure and before being discharged home. The enrollment delay was designed to allow patient recovery from the procedure that would allow for sufficient awareness after the procedure and reduce potential effects of sedation on the readiness to learn. As the study was a pilot, and owing to limited funds, a power analysis was not conducted. After enrollment, a study coordinator collected baseline demographic details. The age and gender of each of the participants were recorded. The races of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asian, or multiracial. For ethnicity, the participants were recorded. The ages of the participants were white, black/African American, American Indian/Alaska Native, Asia
and the percentage of participants with assistance was determined.

The following psychometric evaluations were administered: Hospital Anxiety and Depression Scale (HADS) [23], Rapid Estimate of Adult Literacy in Medicine-Short Form (REALM-SF) [24], Burden-Benefit questionnaire [25], 36-Item Short Form Health Survey (SF-36) [26], and PCI knowledge questionnaire [27] (Multimedia Appendix 1). The REALM-SF scored literacy using grade levels from below a third-grade reading level to a high school reading level and above. HADS scored anxiety and depression by summing the values of the anxiety and depression answers. The SF-36 scores were divided into 6 sections—physical functioning, role limitations because of physical health and because of emotional problems, energy and fatigue, emotional well-being, social functioning, pain, and general health—and the scores were determined by summing weighted values from the answers to the questionnaire and then taking the average of those values for each subscore. The PCI knowledge questionnaire scored participants on the accuracy of their answers; the values were summed to determine the knowledge scores. Finally, the burden benefit questionnaire recorded the percentage of participants that found the study burdensome or beneficial and why. Additional information about household members and friends helping with medication pickup was also included. The goal was to evaluate the differences between the control group and intervention group. Multiple forms of questions helped determine all of the pharmacies where the participants filled their prescriptions. This information was utilized to determine when and if they filled their DAPT prescriptions.

Upon completion of the surveys, the research participants were randomized using block randomization as part of the clinical trial management system to either the control arm (usual care) or the interventional arm (MyIDEA supplementation). The participant in the interventional arm had the MyIDEA program customized to their PCI procedure results. MyIDEA was a customized educational tablet with patient narratives (short stories about patients with the common reason for DAPT discontinuation) to educate them about the importance of DAPT [21]. Full details of patient participatory design are given elsewhere [21]. The control arm participants were given the opportunity to play checkers or tic-tac-toe to control for the novelty of interacting with a tablet. The control arm had traditional educational material that included commercially prepared material. The control group were not asked questions about the material they received, as the study could have approached them before the formal patient education. The nurse that gave the participant the tablet occasionally helped the participant advance the app but was not involved in the intervention. There were no cointerventions. After interacting with the tablet, all patients had their blood drawn to collect the VerifyNow aspirin reactive units (ARU) and P2Y12 reactive units (PRU) [28]. During the PCI procedure, the patients were administered aspirin and a P2Y12 inhibitor (clopidogrel/ticagrelor/prasugrel). These tests were used to measure platelet suppression resulting from the medication. The medication fill rate was collected for a duration of 3 months to calculate pharmacy days covered (PDC).

All research participants were scheduled for a second encounter with the research team during their follow-up cardiology appointment 1 to 2 weeks later. During this second research encounter, the participants first completed the Morisky 8-item medication adherence questionnaire [29], the PCI knowledge questionnaire, and SF-36. After the questionnaires, they interacted with the tablet. Coronary disease–related outcomes such as repeat coronary intervention or myocardial infarction were determined. During this visit, another blood draw was done for the VerifyNow ARU and PRU. The interventional participants interacted with MyIDEA a second time, and the control participants were encouraged to play games on the tablet.

A third appointment was scheduled at 3 months after the initial hospitalization. The same instruments from the second visit were also completed at that visit. One final blood draw was conducted to collect samples for VerifyNow ARU and PRU. Routine phone calls and letters were sent to the participant to encourage follow-up with the research team. All participants who did not attend their visit received at least three follow-up phone calls to reschedule appointments.

For the participants randomized to the interventional arm, additional data were collected during their usage of the MyIDEA program. While using MyIDEA, participants were asked questions, and their responses were recorded using the tablet microphone. The responses reflected the participants’ learning in their module and the symptoms before the procedure. In addition, time stamps of each click or change in the program were recorded. Five patient stories were weaved throughout the program; participants were asked to record which patient narrative they most related to. No outcomes were changed throughout the study.

**Statistical Analysis**

Although the data were extracted and compiled, the research assistants who did not consent the participants and the principal investigator-maintained blinding of the participants. The first 24 participants were unblinded for an intermediate evaluation [27]. However, this study is a complete analysis of all 45 participants. The research team that contacted individuals for recruitment, calculated the PDC, and collected analytics were blinded. However, the nurses that consented participants, the ones who gave the participants the tablet, and the participants were unblinded. The group assignment for all participants was unblinded when the biostatistician calculated the results. The multidisciplinary team and patient advisors reviewed the results and analysis. Descriptive statistics for outcomes and patient characteristics were calculated. A regression analysis of the VerifyNow ARU and PRU was conducted to compare the 2 groups. An imputation of the missing ARU and PRU was conducted via a multivariate imputation by chained by R 3.4.1 (R Foundation)[30]. For other group comparisons, an independent two-tailed t test, a Fisher test, a Mann-Whitney test, or a chi-square test was applied.

To examine the correlation between the level of engagement and the time spent responding to questions in the interventional group, recordings from the participants were transcribed by OE and analyzed. Upon completing the transcription, each participant was labeled as engaged or unengaged.
Engagement was determined by using the tone of the participant’s voice and the total length of recording, which could include background noise, nurses, and pauses. More specifically, tone was described as being attentive or disregarding. Participant speaking time was calculated using Adobe Audition, isolating the voice of the participant for the time calculation. The total time the participant spent talking on MyIDEA was calculated for each visit. The average time each participant spent talking on the app was averaged across visits. To test for significance, a Mann-Whitney test was performed using a 0.05 significance level. A secondary analysis looked at engaged vs unengaged and the PCI knowledge questionnaire.

**Results**

A total of 287 potential research participants were eligible to be recruited during the study at the 2 hospitals (Figure 1). A total of 131 participants did not meet the inclusion criteria, as many of the patients spoke Spanish, and 18 of the potential participants were younger than 50 years. A total of 43 potential participants were missed (ie, the research team was notified but did not approach the participant). For example, 33 potential participants from Cook County had a PCI on the weekend when no research staff was available to recruit.

A total of 45 participants were recruited to the study. One participant withdrew during the first encounter. A total of 23 participants were allocated to the intervention, and 21 participants were allocated to the control. Of the 23 participants randomized to the intervention, 22 participants were given the intervention, as 1 participant left the hospital before the

---

**Figure 1.** Flow diagram of research participants.
information was provided. In the intervention group, 15 participants attended the second visit, and 13 participants attended the third visit. Of the 21 randomized to the control group, 11 attended the second visit, and 11 attended the third visit. A total of 44 research participants had data from pharmacy refills, as 1 participant withdrew. One participant died in the control group. In 17 blood draws out of the potential 131 blood draws, there was a processing error in both or one of the blood draws. The most common reason was a processing error in the lab (11 times; see Figure 1)

The preintervention demographics are outlined in Table 1. The mean age of the participants was 60.4 years (SD 7.1). The mean ages between groups were not significantly different. Of the participants, 60% (27/45) were men, and 40% (18/45) were women (see Table 1). In the control group, 68% (15/24) were men and 32% (7/24) were women, whereas in the MyIDEA group, 48% (11/21) were men and 52% (12/21) were women, but the difference was not statistically significant. The participants were 27% (12/45) white, 60% (27/45) black, 7% (3/45) Asian, and 7% (3/45) multiracial and 7% (3/35) Hispanic. For both groups, the majority of the patients were black, with white being the second largest group, and the rest of the patients being evenly split between Asian and multiple races. The grade reading level of the participants was 2% (1/45) for less than third grade, 9% (4/45) for between fourth and sixth grade, 47% (21/45) for between seventh and eighth grade, and 42% (19/45) for ninth grade and above per REALM-SF. For both the control and MyIDEA groups, the majority of patients’ reading level was seventh grade or above. For medication pickup assistance, 78% of patients (77% of control and 78% of intervention groups) had friends or family who could help pick up prescriptions. The individuals who would assist in medication pickup also benefited from the basic instructions for better conformance to postprocedure instructions from MyIDEA if they were present during the study.
Table 1. Preintervention data for all participant, control, and interventional (My Interventional Drug-Eluting Stent Educational App) groups.

<table>
<thead>
<tr>
<th>Data category</th>
<th>All</th>
<th>Control</th>
<th>MyIDEA³</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>60.4 (7.0)</td>
<td>60.9 (6.9)</td>
<td>60.0 (7.1)</td>
<td>.67</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>27 (60)</td>
<td>15 (68)</td>
<td>11 (48)</td>
<td>.28</td>
</tr>
<tr>
<td>Female</td>
<td>18 (40)</td>
<td>7 (32)</td>
<td>12 (52)</td>
<td></td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.88</td>
</tr>
<tr>
<td>White</td>
<td>12 (27)</td>
<td>6 (27)</td>
<td>6 (26)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>27 (60)</td>
<td>14 (64)</td>
<td>13 (57)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>3 (7)</td>
<td>1 (5)</td>
<td>2 (9)</td>
<td></td>
</tr>
<tr>
<td>Multiple races</td>
<td>3 (7)</td>
<td>1 (5)</td>
<td>2 (9)</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.53</td>
</tr>
<tr>
<td>Hispanic</td>
<td>3 (7)</td>
<td>2 (9)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>42 (93)</td>
<td>20 (91)</td>
<td>22 (96)</td>
<td></td>
</tr>
<tr>
<td><strong>Medication pickup assistance, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients who had assistance picking up medication</td>
<td>35 (78)</td>
<td>17 (77)</td>
<td>18 (78)</td>
<td>.94</td>
</tr>
<tr>
<td><strong>Rapid Estimate of Adult Literacy in Medicine-Short Form n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.71</td>
</tr>
<tr>
<td>&lt;3</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>4-6</td>
<td>4 (9)</td>
<td>3 (14)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>7-8</td>
<td>21 (47)</td>
<td>11 (50)</td>
<td>10 (43)</td>
<td></td>
</tr>
<tr>
<td>9+</td>
<td>19 (42)</td>
<td>8 (36)</td>
<td>11 (49)</td>
<td></td>
</tr>
<tr>
<td><strong>36-Item Short Form Health Survey, value on the test (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>52.7 (30.3)</td>
<td>47.3 (30.3)</td>
<td>57.8 (30.1)</td>
<td>.26</td>
</tr>
<tr>
<td>Role limitations because of physical health</td>
<td>35.2 (40.5)</td>
<td>30.0 (40.2)</td>
<td>40.2 (41.1)</td>
<td>.41</td>
</tr>
<tr>
<td>Role limitations because of emotional problems</td>
<td>57.8 (45.0)</td>
<td>56.1 (47.3)</td>
<td>59.4 (43.8)</td>
<td>.81</td>
</tr>
<tr>
<td>Energy and fatigue</td>
<td>46.0 (19.8)</td>
<td>42.5 (18.8)</td>
<td>49.3 (20.5)</td>
<td>.26</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>74.3 (19.3)</td>
<td>70.4 (20.5)</td>
<td>78.1 (17.6)</td>
<td>.19</td>
</tr>
<tr>
<td>Social functioning</td>
<td>63.3 (29.6)</td>
<td>62.4 (27.6)</td>
<td>64.1 (32.0)</td>
<td>.85</td>
</tr>
<tr>
<td>Pain</td>
<td>54.0 (29.4)</td>
<td>51.2 (31.8)</td>
<td>56.7 (27.2)</td>
<td>.54</td>
</tr>
<tr>
<td>General health</td>
<td>52.4 (19.2)</td>
<td>50.7 (18.0)</td>
<td>53.9 (20.5)</td>
<td>.59</td>
</tr>
<tr>
<td><strong>Hospital Anxiety and Depression Scale, value on the test (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>5.8 (4.5)</td>
<td>7.3 (4.5)</td>
<td>4.3 (4.0)</td>
<td>.03</td>
</tr>
<tr>
<td>Depression</td>
<td>4.5 (3.1)</td>
<td>5.3 (3.3)</td>
<td>3.8 (2.7)</td>
<td>.11</td>
</tr>
<tr>
<td><strong>Burden-Benefit, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Was participation burdensome to you in any way?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>30 (68)</td>
<td>16 (76)</td>
<td>14 (61)</td>
<td>N/A³b</td>
</tr>
<tr>
<td>A little bit</td>
<td>7 (16)</td>
<td>1 (5)</td>
<td>6 (26)</td>
<td>N/A</td>
</tr>
<tr>
<td>Somewhat</td>
<td>5 (11)</td>
<td>3 (14)</td>
<td>2 (9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Quite a bit</td>
<td>1 (2)</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Very much</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>No answer</td>
<td>1 (2)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Was participation beneficial to you in any way?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all</td>
<td>3 (7)</td>
<td>2 (10)</td>
<td>1 (4)</td>
<td>N/A</td>
</tr>
</tbody>
</table>
The HADS anxiety and depression scores for MyIDEA (mean 4.3, SD 4 anxiety and mean 3.8, SD 2.7 depression) and control groups (mean 7.3, SD 4.5 anxiety and mean 5.3, SD 3.3 depression) were statistically significantly different for anxiety. HADS has scores ranging from 0 to 21, where 0 to 7 is normal, 8 to 10 is borderline, and 11 to 21 is abnormal before the intervention. The depression and anxiety were only measured on enrollment. The group differences in average PCI knowledge questionnaire scores or in Morisky medication adherence belief scores for both visits were not statistically significant.

The majority of participants in both the MyIDEA and control groups did not find the study burdensome (65% of MyIDEA and 76% of control groups). However, the two major reasons that the study was burdensome to some patients were that either the interview was too long or there were too many questions and the patient was too weak or ill at times. Similarly, the majority of participants in both the MyIDEA and control groups found participation in the study at least somewhat beneficial. The top two reasons for which they found the study beneficial were that it helped them to think about these topics and that it made them feel good to help others or contribute to society.

SF-36 was given at the first visit, and scores were reported within the normal ranges (see Table 1 and 2).

A comparison of MyIDEA and control groups for medication adherence showed no significant difference in ARU, PRU, or PDC (Tables 3 and 4). The prescription pattern changed midstudy from 30-day prescriptions to 90-day prescriptions. The overall well-being throughout the study was evaluated by using SF-36. The group differences in various domains including energy and fatigue, emotional well-being, social functioning, pain, and general health were not statistically significant. There was a little change from the hospital to the second and third visits.

For the patients in the MyIDEA group, the average amount of time spent on the app in the first visit was 17.2 min (SD 4.4) and in the second visit was 13.8 min (SD 5.5). The average percentage of time spent on the slides was lowest for the tutorial and chapter introduction slides (Figure 2). The selection of patient narratives is reported in Table 5.
### Table 2. Postintervention variables.

<table>
<thead>
<tr>
<th>Postintervention variable</th>
<th>All scoring unit, mean (SD)</th>
<th>Control scoring unit, mean (SD)</th>
<th>Interventional scoring unit, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visit 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morisky</td>
<td>2.4 (0.7)</td>
<td>2.5 (0.6)</td>
<td>2.3 (0.7)</td>
<td>.50</td>
</tr>
<tr>
<td>Percutaneous coronary in-</td>
<td>7.9 (1.6)</td>
<td>7.7 (1.5)</td>
<td>8.0 (1.7)</td>
<td>.52</td>
</tr>
<tr>
<td>tervention knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>36-Item Short Form Health Survey</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>55.9 (30.4)</td>
<td>53.5 (30.3)</td>
<td>58.2 (30.2)</td>
<td>N/A^a</td>
</tr>
<tr>
<td>Role limitations due to physical health</td>
<td>43.5 (40.0)</td>
<td>38.5 (40.0)</td>
<td>48.2 (39.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>Role limitations due to emotional problems</td>
<td>60.5 (42.6)</td>
<td>53.8 (40.3)</td>
<td>66.7 (43.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Energy and fatigue</td>
<td>54.3 (42.6)</td>
<td>52.3 (22.4)</td>
<td>56.1 (19.4)</td>
<td>N/A</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>54.3 (21.0)</td>
<td>70.8 (17.8)</td>
<td>86.9 (12.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Social functioning</td>
<td>76.9 (24.0)</td>
<td>72.1 (21.5)</td>
<td>81.3 (25.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pain</td>
<td>66.3 (29.8)</td>
<td>63.7 (28.5)</td>
<td>68.8 (30.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>General health</td>
<td>58.9 (18.5)</td>
<td>57.7 (19.6)</td>
<td>60.0 (17.3)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Visit 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morisky</td>
<td>2.5 (1.4)</td>
<td>2.5 (1.3)</td>
<td>2.5 (1.6)</td>
<td>.94</td>
</tr>
<tr>
<td><strong>36-Item Short Form Health Survey</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>52.0 (27.8)</td>
<td>31.7 (23.3)</td>
<td>68.6 (20.9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Role limitations due to physical health</td>
<td>45.0 (46.5)</td>
<td>22.2 (44.1)</td>
<td>63.6 (43.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Role limitations due to emotional problems</td>
<td>65.0 (44.1)</td>
<td>63.0 (45.5)</td>
<td>66.7 (47.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Energy and fatigue</td>
<td>53.8 (18.8)</td>
<td>50.6 (20.2)</td>
<td>56.4 (19.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Emotional well-being</td>
<td>75.4 (14.7)</td>
<td>73.8 (15.1)</td>
<td>76.7 (15.6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Social functioning</td>
<td>75.6 (25.1)</td>
<td>72.2 (29.2)</td>
<td>78.4 (23.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pain</td>
<td>62.0 (28.4)</td>
<td>51.7 (32.2)</td>
<td>70.5 (24.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>General health</td>
<td>52.0 (12.8)</td>
<td>43.9 (11.7)</td>
<td>58.6 (10.5)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Proportion of days covered (PDC)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PDC &gt;0.8</td>
<td>N/A</td>
<td>14 (78)</td>
<td>16 (80)</td>
<td>.99</td>
</tr>
<tr>
<td>PDC ≤0.8</td>
<td>N/A</td>
<td>4 (22)</td>
<td>4 (20)</td>
<td>N/A</td>
</tr>
<tr>
<td>Adverse events (participants)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Death</td>
<td>N/A</td>
<td>1</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Rehospitalizations all</td>
<td>N/A</td>
<td>8</td>
<td>11 (9)</td>
<td>N/A</td>
</tr>
<tr>
<td>Unplanned heart-relat-</td>
<td>N/A</td>
<td>2</td>
<td>4 (3)</td>
<td>N/A</td>
</tr>
<tr>
<td>ed hospitalizations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^aN/A: Not applicable.
Table 3. Aspirin reactive units and P2Y12 reactive units (PRU) with tablet data. Linear regression of aspirin reactive units (ARU) and PRU.

<table>
<thead>
<tr>
<th>Dependent variable and predictor</th>
<th>Estimate</th>
<th>SE</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARU(^{a})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group (reference: control)</td>
<td>10.55</td>
<td>22.56</td>
<td>0.47</td>
<td>.64</td>
</tr>
<tr>
<td>Baseline ARU</td>
<td>0.09</td>
<td>0.16</td>
<td>0.58</td>
<td>.57</td>
</tr>
<tr>
<td>PRU(^{b})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group (reference: control)</td>
<td>−9.42</td>
<td>22.39</td>
<td>−0.42</td>
<td>.68</td>
</tr>
<tr>
<td>Baseline PRU</td>
<td>0.52</td>
<td>0.13</td>
<td>3.9</td>
<td>.001</td>
</tr>
</tbody>
</table>

\(^{a}\)ARU: aspirin reactive units.

\(^{b}\)PRU: P2Y12 reactive units.

Table 4. The average time the research subjects spoke when recording answers to questions sorted by engaged and unengaged participants.

<table>
<thead>
<tr>
<th>Average time on app</th>
<th>Engaged (seconds)</th>
<th>Unengaged (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample SD</td>
<td>9.35</td>
<td>2.01</td>
</tr>
<tr>
<td>Variance</td>
<td>87.41</td>
<td>4.03</td>
</tr>
<tr>
<td>Total number</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Sum</td>
<td>111.86</td>
<td>44.29</td>
</tr>
<tr>
<td>Mean</td>
<td>12.43</td>
<td>3.41</td>
</tr>
</tbody>
</table>

Figure 2. The average percentage of time spent on slides.
A participant who was attentive spent, on average, a longer time on the app than a participant who was disengaged. The length of time a research participant spent on each slide was associated with the tone and content, as determined by the trained listener. It was determined that a participant labeled as engaged spoke for an average of 12.4 seconds per question, whereas an unengaged patient spoke an average of 3.4 seconds per question. The Mann-Whitney test yielded a value $P<.001$. We compared the engaged vs the unengaged for the reading grade level of both groups. A t test comparing the scores gives $P=.09$. A Fisher test of proportion of high school graduates between the 2 groups is $P=.33$. The average PCI score for engaged participants was 8.3 and for the unengaged was 7.6.

**Discussion**

**Principal Findings**

Adherence to DAPT is a key element to optimal prevention of negative outcomes post DES placement. MyIDEA demonstrated the feasibility of elderly patients (>50 years) who were able and willing to learn via a tablet device. There was no significant difference between the groups DAPT and PDC because of the small sample size. However, this test only measures recent medication adherence. In this study, PDC was confounded because of the change in practice from monthly prescriptions to 3-month prescriptions for the P2Y12 inhibitors. This change increased the participants’ PDC.

Elderly patients are willing to use a tablet device to increase their knowledge about medication adherence. Additional studies with larger samples are needed to measure the true clinical impact of novel educational interventions. As technology continues to advance, the challenge of transforming detailed clinical information into a form digestible by patients will continue to need innovation. The difference in time of the respondents between the engaged and the unengaged could be used to encourage health practitioners to identify patients in need of more education or a diversity of education. As this study had a challenge recruiting research subjects, all of the perspectives and potential implications of a hospital-wide deployment of a similar educational program would be challenging to predict. Furthermore, very few subjects found the study burdensome, but the potential benefit was a more nuanced answer with 20% saying “a little bit to not at all” for benefits.

**Intriguing Findings**

One surprising finding was the knowledge between the 2 groups was not statistically different in the 2 groups on the second visit from the PCI knowledge questionnaire. The average for the second visit in the control group was 7.7 and the intervention group was 8.0, with the baseline average from the first visit being 7.9. A number of reasons could cause this observation. One is a ceiling effect where you can only score 10 as the highest on the questionnaire. Second, the questionnaire was administered 2 to 4 weeks after the intervention but before the second interaction with the tablet. This was a test of the recall from weeks ago and not immediate recall. In addition, as the MyIDEA educational program focused on both factual information as well as patient narratives about the impact of missing medication, the quiz only focused on factual knowledge. Future knowledge quizzes should focus on applied knowledge as well to measure the complete impact of this type of education.

The MyIDEA and control groups’ SF-36 scores were within the average SF-36 scores for patients with ischemic heart disease (IHD) in the United States [31]. The physical functioning scores for the intervention (57.8) and control (47.3) groups were within the average score of US patients with IHD (58.8, SD 27.4). Similarly, the role limitations due to physical health scores for the intervention (40.2) and control (30) groups were within the range of the average score for US patients with IHD (45.0, SD 42.7). The average energy and fatigue scores of the intervention (49.3) and control (42.4) groups were also within the range of patients with IHD (52.8, SD 20.4).

Emotional well-being scores for the intervention (78.1) and control (71.0) groups were within the range of patients with IHD (73.1, SD 17.9). Social functioning scores for the MyIDEA group (64.1) and the control group (62.5) were also within the range of average scores of patients with IHD (76.0, SD 23.8). The average scores for pain for the interventional (56.7) and control (50.5) groups were also within the average scores for patients with IHD (64.4, SD 25.0). Finally, the average scores for general health for the MyIDEA group (53.9) and the control group (50.5) were within the range for US patients with IHD (54.7, SD 22.2).

**Limitations**

The limitations of the study include having only 2 urban hospitals to recruit the research participants. The number of participants recruited was not sufficient for a properly powered study but instead to evaluate feasibility of recruitment and use of the tablet. As tablets and phones have become more interchangeable than when this project first began, a portable version of the program should be designed for future studies. The strength of this study was the ability of the elderly to use a customized tablet program both in the hospital and the

<table>
<thead>
<tr>
<th>Patient narrative name</th>
<th>First interaction (number of subjects selecting)</th>
<th>Second interaction (number of subjects selecting)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blank</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Frank</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Eva</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Heather</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Joanne</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5. Patient selection of most similar to during My Interventional Drug-Eluting Stent Educational App program.
outpatient clinic, opening up new possibilities in forms of education.

**Comparison With Commercial Products**

Several mobile health apps exist, but these products simply attempt to measure medication adherence. For example, GlowCap reminds patients to take their medication by glowing, sounding an alarm, and calling their homes [32]. However, because the patient only needs to open the medicine bottle to have their adherence recorded, it is not an accurate measurement of adherence. Other apps do not target the reasons behind nonadherence, such as a lack of knowledge about the effects of nonadherence.

MyIDEA was created to intervene before the patient begins to take their medication, whereas other solutions target the patient as they begin their medication. For example, one intervention delivered customized SMS text messages at the time patients were supposed to take their medication for 30 days [32]. This app provides educational reasons to take medicine as well as reminders. However, there is no indication that those patients would always receive the SMS text message or that the patients took the medication. As a result, the educational engagement of this intervention is limited. Another approach to solving adherence is the use of smartphone apps such as MoviPill, a mobile game that connects elderly participants to a social network and awards points for taking their medication close to the prescribed time [33]. This app utilizes social rewards to persuade the patient to take their medication at the right time. However, it only targeted the patients interested in playing the game and did not educate them on why they should take their medicine. Thus, MyIDEA is unique because it begins to educate patients about the importance of medication adherence before they begin to take their medicine and continues the education weeks after the procedure.

As typical with mobile health interventions, the participants were unblinded. Future research could be conducted to compare the level of engagement with medication adherence as well as by working with the clinical team in ensuring patients engage in their own care. If there was a correlation between medication adherence and the level of engagement, MyIDEA could be an effective tool for nurses and other health care providers to better understand the needs of their patients and increase medication adherence.

**Acknowledgments**

This project was funded by the Department of Biomedical and Health Information Sciences and the Center of Excellence for End-of-Life Transition Research. The project described was supported by Grant Number P30AG022849 from the National Institute on Aging, Grant Number UL1 TR000050 from the National Center for Advancing Translational Sciences, and Grant P30 NR010680 from the National Institute of Nursing Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging, National Center for Advancing Translational Sciences, National Institute of Nursing Research, or the National Institutes of Health. We would like to thank Teresa Dominguez for the discussion and analysis of the findings from this project.

**Conflicts of Interest**

None declared.

**Editorial notice:** This randomized study was not prospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1

PCI Knowledge Questionnaire.

[PDF File (Adobe PDF File), 94 KB - mhealth_v8i6e15900_app1.pdf]

Multimedia Appendix 2

CONSORT e-HEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 488 KB - mhealth_v8i6e15900_app2.pdf]

**References**


Abbreviations

- ARU: aspirin reactive units
- DAPT: dual antiplatelet therapy
- DES: drug-eluting stent
- HADS: Hospital Anxiety and Depression Scale
- IHD: ischemic heart disease
- MyIDEA: My Interventional Drug-Eluting Stent Educational App
- PCI: percutaneous coronary intervention
- PDC: proportion of days covered
- PRU: P2Y12 reactive units
- REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form
- SF-36: 36-Item Short Form Health Survey
Influence of the SARS-CoV-2 Outbreak on the Uptake of a Popular Smoking Cessation App in UK Smokers: Interrupted Time Series Analysis

Olga Perski¹, PhD; Aleksandra Herbeć¹,², PhD; Lion Shahab¹, PhD; Jamie Brown¹, PhD

¹Department of Behavioural Science and Health, University College London, London, United Kingdom
²Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Olga Perski, PhD
Department of Behavioural Science and Health
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 020 7679 1258
Email: olga.perski@ucl.ac.uk

Abstract

Background: The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) outbreak may motivate smokers to attempt to stop in greater numbers. However, given the temporary closure of UK stop smoking services and vape shops, smokers attempting to quit may instead seek out mobile health support, such as smartphone apps.

Objective: We examined, using an interrupted time series approach, whether the SARS-CoV-2 outbreak has been associated with a step change or increasing trend in UK downloads of an otherwise popular smoking cessation app, Smoke Free.

Methods: Data were from daily and nondaily adult smokers in the United Kingdom who had downloaded the Smoke Free app between January 1, 2020, and March 31, 2020 (primary analysis), and January 1, 2019, and March 31, 2020 (secondary analysis). The outcome variable was the number of downloads aggregated at the 12-hourly (primary analysis) or daily level (secondary analysis). The explanatory variable was the start of the SARS-CoV-2 outbreak, operationalized as March 1, 2020 (primary analysis), and January 15, 2020 (secondary analysis). Generalized additive mixed models adjusted for relevant covariates were fitted.

Results: Data were collected on 45,105 (primary analysis) and 119,881 (secondary analysis) users. In both analyses, there was no evidence for a step change or increasing trend in downloads attributable to the start of the SARS-CoV-2 outbreak. Calculation of Bayes factors (BFs) indicated that the data for the primary analysis favored the null hypothesis compared with large associations (for level, BF=0.25; for slope, BF=0.26) but were insensitive to the detection of small associations (for level, BF=0.78; for slope, BF=1.35).

Conclusions: In the United Kingdom, between January 1, 2020, and March 31, 2020, and between January 1, 2019, and March 31, 2020, there was no evidence that the SARS-CoV-2 outbreak has been associated with a large step change or increasing trend in downloads of a popular smoking cessation app. Findings on the association of the SARS-CoV-2 outbreak with a small step change or increasing trend were inconclusive.

(JMIR Mhealth Uhealth 2020;8(6):e19494) doi:10.2196/19494

KEYWORDS
SARS-CoV-2; COVID-19; smoking cessation; mobile health; smartphone app; time series analysis; smoking; public health; app

Introduction

The global outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes the coronavirus disease (COVID-19) [1], was imported into the United Kingdom in January 2020 [2]. Community transmission has subsequently been established [3]. The growing spread of SARS-CoV-2 has been accompanied by a series of government announcements of escalating social distancing policies and legislation. On March
8, 2020, prime minister Boris Johnson unveiled the UK Coronavirus Action Plan [4], which was followed by government advice to practice social distancing on March 16 and behavioral restrictions enforceable by law on March 23, 2020.

Definitive evidence on whether current smokers are at increased risk of disease, morbidity, and mortality from COVID-19 are not yet available. However, researchers and clinicians have emphasized biological (eg, reduced respiratory immune defense) and behavioral (eg, repetitive hand-to-mouth movements) factors that may mean that smokers are at an increased risk of contracting SARS-CoV-2 [5-7], and early evidence indicates that hospitalized smokers may be more likely to experience complications and mortality from COVID-19 compared with nonsmokers [8,9]. UK public health messaging focused on advice to stop smoking has been limited. On March 20, 2020, the charitable organization Action on Smoking and Health launched an online stop smoking campaign labelled “Today Is The Day” [10] with an accompanying hashtag, #QuitForCovid (see Table 1 for a timeline of key UK events related to smoking and COVID-19). On April 3, 2020, Public Health England released a news story headlined “Smokers at greater risk of severe respiratory disease from COVID-19” [11], which generated substantial coverage [12,13].

Table 1. Timeline of key UK events in 2020 related to smoking and COVID-19, including government policies and legislation [14].

<table>
<thead>
<tr>
<th>Date</th>
<th>Key event</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 31</td>
<td>First two cases reported in the United Kingdom</td>
</tr>
<tr>
<td>February 7</td>
<td>Primary care guidance updated to state that the virus is most likely to be seen in travelers returning from China, Hong Kong, Japan, Macau, Malaysia, Republic of Korea, Singapore, Taiwan, or Thailand</td>
</tr>
<tr>
<td>February 21</td>
<td>At least 17 temporary GP(^a) closures occur</td>
</tr>
<tr>
<td>February 24</td>
<td>4 UK passengers repatriated from a Japanese cruise ship test positive for COVID-19(^b), bringing the total number of UK cases to 13</td>
</tr>
<tr>
<td>March 2</td>
<td>GP practice closures continue</td>
</tr>
<tr>
<td>March 6</td>
<td>First UK patient dies from COVID-19</td>
</tr>
<tr>
<td>March 8</td>
<td>Prime Minister Boris Johnson unveils the Coronavirus Action Plan</td>
</tr>
<tr>
<td>March 11</td>
<td>World Health Organization declares the severe acute respiratory syndrome coronavirus 2 outbreak a pandemic</td>
</tr>
<tr>
<td>March 12</td>
<td>UK COVID-19 risk level raised from moderate to high</td>
</tr>
<tr>
<td>March 16</td>
<td>UK government advises on social distancing (eg, avoiding nonessential travel and contact with others); pregnant women, adults 70 years or older, and individuals with health conditions are urged to self-isolate</td>
</tr>
<tr>
<td>March 18</td>
<td>UK government announces that schools will close to most students, with children of key workers and vulnerable children still able to attend</td>
</tr>
<tr>
<td>March 20</td>
<td>Restaurants, pubs, clubs, and indoor sport and leisure centers ordered to close; Action on Smoking and Health launches the stop smoking campaign “Today Is The Day” and #QuitForCovid</td>
</tr>
<tr>
<td>March 23</td>
<td>UK government announces restrictions made on freedom of movement, enforceable by law</td>
</tr>
<tr>
<td>April 3</td>
<td>Public Health England releases a news story emphasizing the increased risk of severe respiratory disease due to COVID-19 in smokers</td>
</tr>
</tbody>
</table>

\(a\)GP: general practice. \\
\(b\)COVID-19: coronavirus disease.

Options for UK smokers interested in quitting usually include specialist stop smoking services and vape shops. However, the introduction of social distancing policies and behavioral restrictions enforceable by law mean that the vast majority of services and shops have temporarily closed. At the same time, it is plausible that the SARS-CoV-2 outbreak has acted as a “teachable moment,” motivating smokers to attempt to stop in greater numbers than would otherwise be observed for this time of the year. Given these temporary closures of stop smoking services and vape shops, smokers attempting to quit during the SARS-CoV-2 outbreak may have to seek out digital support, such as websites and smartphone apps. If so, it would be important for public health bodies to put resources toward ramping up access to high-quality, evidence-informed digital support. One way of assessing whether the recent SARS-CoV-2 outbreak has been accompanied by a surge in smokers accessing digital support is to use an interrupted time series approach to examine whether there has been a step change or increasing trend in downloads of an otherwise popular smoking cessation app. The Smoke Free app, designed for English speakers, includes behavior change techniques that can improve the chances of quitting and shows early evidence of effectiveness [15]. The app is live on the Apple and Google Play stores and has a large user base with >1 million global downloads per annum. The app exists as a free version with a set of core features and a paid (“pro”) version with additional content.

This study used an interrupted time series approach to assess whether the SARS-CoV-2 outbreak up to the end of March 2020 (ie, a hypothesized “teachable moment”) has been accompanied
by an increase in UK downloads of the free and “pro” versions of the Smoke Free app (henceforth referred to as the Smoke Free app). Two different starting points for the outbreak were used. First, we assessed whether there has been a step change or increasing trend (linear or nonlinear) in downloads during a period of government announcements of escalating social distancing policies and legislation, with March 1, 2020, as the stipulated starting point (see Table 1; a series of government announcements occurred throughout March 2020). Second, we assessed whether there has been an increasing trend (linear or nonlinear) in downloads over the course of the SARS-CoV-2 outbreak so far, compared with the preceding 12 months, with January 15, 2020, as the stipulated starting point. The middle of January 2020 was selected as a starting point to reduce confounding by the regularly observed annual spike in downloads from late December to mid-January.

Specifically, we aimed to address the following research questions:

1. Over a 3-month time series (ie, January 1, 2020, to March 31, 2020), has a period of government announcements of escalating social distancing policies and legislation (ie, March 1, 2020, to March 31, 2020) been accompanied by a step change or increasing trend in UK downloads of the Smoke Free app?

2. Over a 15-month time series (ie, January 1, 2019, to March 31, 2020), has the SARS-CoV-2 outbreak thus far (ie, January 15, 2020, to March 31, 2020) been accompanied by an increasing trend in UK downloads of the Smoke Free app?

Methods

Study Design

This was a natural experiment without active recruitment. The study protocol and analysis plan were preregistered on the Open Science Framework [16]. The study involved the analysis of anonymized app data at the aggregate level and fell within the scope of “The optimisation and implementation of interventions to change behaviours related to health and the environment,” approved by University College London’s Research Ethics Committee (Project ID: CEHP/2020/579). Users provide consent for their data to be used in research by virtue of downloading the app and agreeing to the terms and conditions.

Study Setting and Population

To be included in the analytic sample, daily and nondaily smokers needed to be 18 years or older and have downloaded the Smoke Free app from a UK-based app store (ie, Google Play and Apple App Store) during the respective study periods.

Measures

Outcomes

For the primary analysis, the outcome of interest was the number of UK Smoke Free app downloads aggregated at the 12-hourly level. This temporal unit was selected to increase the number of observations in the postintervention period.

For the secondary analysis, the outcome of interest was the number of UK Smoke Free app downloads aggregated at the daily level.

Exposures

For the primary analysis, the 60-day period (or 120 observations) before the government announcements of escalating social distancing policies and legislation was coded as 0, and the 31-day period (or 62 observations) containing the announcements was coded as 1. The trend during the 31-day postintervention period from March 1, 2020, to March 31, 2020, (or 62 observations) was coded from 1-62.

For the secondary analysis, the 379-day period (or 379 observations) before the start of the SARS-CoV-2 outbreak was coded as 0, and the trend over the 77-day period containing the outbreak from January 15, 2020, to March 31, 2020, (77 observations) was coded from 1-77.

Covariates

For the primary analysis, the time of day was included as a categorical covariate (midnight to 11:59 AM or “morning” vs noon to 11:59 PM or “evening”), and the day of the week was included as a cyclic cubic spline to capture nonlinear, seasonal patterns in downloads.

For the secondary analysis, the month of the year and the day of the week were included as cyclic cubic splines to capture nonlinear, seasonal patterns in downloads. Two additional covariates were included. On December 19, 2019, a new expert feature, which involved the ability to communicate directly with stop smoking counsellors, was launched for “pro” users (ie, “expert feature launch”). On December 27, 2019, a promotional offer to purchase a separate version of the app called Smoke Free Plus in a range of Boots stores across the United Kingdom was announced (ie, “national advertising campaign”). The periods before and after the expert feature launch and the national advertising campaign were coded as 0 and 1, respectively.

Data Analysis

The analyses were conducted in R v.3.6.3 (R Foundation for Statistical Computing) using the mgcv package.

For the primary analysis, an interrupted time series analysis (segmented regression) was conducted using a generalized additive mixed model (GAMM). GAMMs take account of seasonality through the fitting of seasonal smoothing terms [17]. First, data were assessed for overdispersion (ie, when the variance is greater than the mean). As there was evidence for overdispersion, a quasi-Poisson distribution was specified. Second, plots of the autocorrelation function and partial autocorrelation function were assessed to identify plausible values for the autocorregressive (AR) and moving average (MA) terms. Third, a segmented regression model was fitted. Different models with plausible AR and MA terms were compared using the Akaike information criterion (AIC), with smaller values indicating better model fit. The final segmented regression model included terms for a secular trend (denoting the number of observations in the entire study period), level (denoting the pre- and postintervention segments), slope (denoting the trend...
in the postintervention segment), time of day, and day of the week. The fit of models including linear, quadratic, and cubic trends for the postintervention segment was assessed using the AIC.

The same steps were repeated for the secondary analysis. The final model was selected in a similar way. As specified in the analytic plan, the model included terms for a secular trend, slope, month of the year, day of the week, the expert feature launch, and the national advertising campaign.

**Unplanned Sensitivity Analyses**

Prompted by inspection of the data and as a result of the review process, three unplanned sensitivity analyses (SAs) were conducted.

First, given uncertainties as to when the SARS-CoV-2 outbreak had started to affect smokers’ lives, we modelled the starting point of the outbreak as approximately 15 days before (ie, February 15, 2020) and after (ie, March 15, 2020) the original starting point (ie, March 1, 2020). These SAs were considered exploratory (as opposed to hypothesis testing).

Second, SAs were conducted to examine whether observed nonsignificant associations for the primary analysis could best be characterized as evidence of no effect or whether the data were insensitive to distinguish the null from the alternative hypothesis [18]. Bayes factors (BFs) with the alternative hypotheses conservatively represented by a half-normal distribution were calculated using an online calculator [19]. With the alternative hypothesis represented by a half-normal distribution, the standard deviation of a distribution can be specified as an expected effect size, meaning that plausible values are represented between zero and twice the effect size, with smaller values being more likely. The expected effect sizes for the level (ie, step change) and slope (ie, the trend in the postintervention period) were set as follows: first, we imagined a large step change on the basis of a hypothetical scenario in which 5% of the 1,890,000 UK smokers who make a quit attempt each year with e-cigarettes or stop smoking services would switch immediately to the Smoke Free app (ie, 94,500/365 = 259). With a base rate of ~400 Smoke Free downloads per day, this would equate to an IRR of 409/400 (IRR=1.02). BFs ≥ 3 can be interpreted as substantial evidence for the alternative hypothesis and against the null. BFs ≤ 1/3 can be interpreted as evidence for the null hypothesis. BFs between 1/3 and 3 suggest that the data are insensitive to distinguish the alternative hypothesis from the null [18].

Third, with additional data becoming available, the primary and secondary analyses were topped up with data up to May 12, 2020. These SAs were considered exploratory.

**Results**

**Analyses**

Data were collected on 45,105 and 119,881 users, respectively. Figures 1 and 2 show graphs of the number of users downloading the app by month of the year for the respective analyses.

Table 2 shows the results for the best fitting segmented regression models for the primary and secondary analysis. In the primary analysis, there was no evidence for a step change or increasing trend in downloads attributable to the start of the SARS-CoV-2 outbreak on March 1, 2020. The rate of downloads was significantly greater in the evenings (IRR=1.780, 95% CI 1.581-2.004, P < .001). In the secondary analysis, there was evidence for a very small but significant, linearly decreasing trend in downloads over the course of the entire study period (IRR=0.996, 95% CI 0.994-0.998, P < .001). There was no evidence for an increasing trend in downloads attributable to the start of the SARS-CoV-2 outbreak on January 15, 2020. There was a substantial step change in downloads following the national advertising campaign (IRR=2.067, 95% CI 1.387-3.079, P < .001).
Figure 1. The number of users downloading the Smoke Free app from January 1, 2020, to March 31, 2020. The dotted blue vertical line represents the stipulated starting point of the severe acute respiratory syndrome coronavirus 2 outbreak (ie, March 1, 2020). The solid red line represents the fitted values with a loess smoothing function applied and associated 95% CIs.

Figure 2. The number of users downloading the Smoke Free app from January 1, 2019, to March 31, 2020. The dotted blue vertical line represents the stipulated starting point of the severe acute respiratory syndrome coronavirus 2 outbreak (ie, January 15, 2020). The solid red line represents the fitted values with a loess smoothing function applied and associated 95% CIs. For clarity, the y-axis was capped at 1300 downloads.
Table 2. Results from the best fitting model for each analysis.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>IRR(^a) (95% CI)</th>
<th>SE</th>
<th>(P) value</th>
<th>Bayes factor (small expected effect)</th>
<th>Bayes factor (large expected effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary analysis: January 1, 2020-March 31, 2020(^b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend (January 1, 2020-March 31, 2020)</td>
<td>0.998 (0.994-1.002)</td>
<td>0.002</td>
<td>.23</td>
<td>N/A(^c)</td>
<td>N/A</td>
</tr>
<tr>
<td>Level (March 1, 2020)</td>
<td>1.142 (0.916-1.423)</td>
<td>0.112</td>
<td>.24</td>
<td>0.78</td>
<td>0.25</td>
</tr>
<tr>
<td>Slope (March 1, 2020-March 31, 2020)</td>
<td>0.999 (0.988-1.010)</td>
<td>0.006</td>
<td>.85</td>
<td>1.35</td>
<td>0.26</td>
</tr>
<tr>
<td>Time of day(^d)</td>
<td>1.780 (1.581-2.004)</td>
<td>0.060</td>
<td>&lt;.001</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Secondary analysis: January 1, 2019-March 31, 2020(^e)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trend (January 1, 2019-March 31, 2020)</td>
<td>0.996 (0.994-0.998)</td>
<td>0.001</td>
<td>&lt;.001</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Slope (January 15, 2020-March 31, 2020)</td>
<td>0.976 (0.908-1.049)</td>
<td>0.037</td>
<td>.51</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Slope(^2)</td>
<td>1.001 (0.998-1.003)</td>
<td>0.001</td>
<td>.52</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Slope(^3)</td>
<td>1.000 (1.000-1.000)</td>
<td>&lt;0.001</td>
<td>.54</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Expert feature launch (December 19, 2019)(^f)</td>
<td>1.385 (0.885-2.168)</td>
<td>0.229</td>
<td>.16</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>National advertising campaign (December 27, 2019)(^f)</td>
<td>2.067 (1.387-3.079)</td>
<td>0.203</td>
<td>&lt;.001</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)IRR: incidence rate ratio.
\(^b\)Adjusted for a second-order autoregressive term and day of the week.
\(^c\)Not applicable.
\(^d\)Referent=morning (vs evening).
\(^e\)Adjusted for a first-order autoregressive term, month of the year, and day of the week.
\(^f\)Modeled as a step change.

Unplanned Sensitivity Analyses

**Alternative Conceptualizations of the Outbreak**

When modeling the starting point of the outbreak as February 15, 2020, and March 15, 2020, there was no evidence for a step change or increasing trend in downloads attributable to the start of the SARS-CoV-2 outbreak (see Multimedia Appendix 1).

**Calculation of Bayes Factors**

The calculation of BFs indicated that the data for the primary analysis favored the null hypothesis compared with large associations (for level, BF=0.25; for slope, BF=0.26) but were insensitive to detection of small associations (for level, BF=0.78; for slope, BF=1.35; see Table 2).

**Analyses With Additional Data**

In the SAs repeating the primary and secondary analyses with additional data up to May 12, 2020, there was no evidence for a step change or increasing trend in downloads attributable to the start of the SARS-CoV-2 outbreak (see Multimedia Appendix 1).

**Discussion**

**Principal Findings**

This study used an interrupted time series approach to examine the impact of the ongoing SARS-CoV-2 outbreak on the uptake of an otherwise popular smoking cessation app in UK smokers. Across two different time periods and using two different conceptualizations of the starting point of the outbreak, we found no evidence for a step change or increasing trend in the number of downloads of the popular Smoke Free app. Calculation of BFs indicated that the data favored the null compared with large associations but were insensitive to detection of small associations.

**Limitations**

First, the small number of observations since the start of the SARS-CoV-2 outbreak means that we likely had low statistical power to detect anything but large effects. Second, pinpointing the exact starting point of the outbreak was not straightforward, as it is unclear when UK smokers first became aware of or affected by the pandemic, hypothesized to act as triggers for additional quit attempts. This was partly mitigated by conducting two SAs, which also did not provide evidence for a step change or increasing trend in downloads. Third, given the current structure of the app’s database, we were unable to distinguish between downloads of the free and “pro” versions of the Smoke Free app in our analyses. Although a tally is kept for free and “pro” downloads combined, the tally kept for “pro” accounts has been designed to track weekly subscriptions (with each user counted several times) as opposed to one-off downloads. It is plausible that a greater number of smokers who otherwise would have tried to stop using e-cigarettes or stop smoking services would now be more inclined to purchase the “pro” version of the app. Fourth, the Smoke Free app is only one of many available online stop smoking platforms in the United Kingdom. We did not capture traffic on other available apps or websites such as the NHS Smokefree website [21] and were, hence, unable to model any effect of the SARS-CoV-2 outbreak on interest in stop smoking support (digital or nondigital) more
broadly. For example, although vape shops are temporarily closed, it is plausible that online sales of e-cigarettes have increased during the study period. Fifth, we did not take account of tobacco industry behavior in our models. There has also been legitimate scientific investigation into the possible links between smoking, nicotine, and SARS-CoV-2 infection and COVID-19 outcomes [22,23]. In the absence of clear evidence as to whether smokers are at increased or reduced risk, it is plausible that the tobacco industry has attempted to capitalize on this uncertainty by, for example, influencing media reports, which may in turn have influenced quitting behavior.

Implications for Policy
The lack of an uptick in downloads of a popular smoking cessation app during the SARS-CoV-2 outbreak thus far suggests that smokers may not be turning to available, evidence-informed digital support in greater numbers. Evidence from controlled studies and population-level surveys indicate that smoking cessation attempts involving pharmacological or behavioral support (including digital interventions) are substantially more likely to be successful compared with unassisted attempts [24-26]. In the absence of readily available alternatives due to social distancing measures, UK public health bodies should consider putting resources toward increasing awareness of available, evidence-informed mobile health support.

Conclusions
In the UK, between January 1, 2020, and March 31, 2020, and between January 1, 2019, and March 31, 2020, there was no evidence that the SARS-CoV-2 outbreak thus far has been associated with a large step change or increasing trend in downloads of a popular smoking cessation app. Findings on the association of the SARS-CoV-2 outbreak with a small step change or increasing trend were inconclusive.

Acknowledgments
OP and JB receive salary support from Cancer Research UK (C1417/A22962). JB, LS, AH, and OP are members of Shaping Public Health Policies to Reduce Inequalities and Harm, a UK Prevention Research Partnership Consortium (MR/S037519/1). UK Prevention Research Partnership is an initiative funded by the UK Research and Innovation Councils, the Department of Health and Social Care (England) and the UK devolved administrations, and leading health research charities.

Authors’ Contributions
OP, AH, LS, and JB designed the study. OP conducted the statistical analyses and wrote the first draft of the manuscript. All authors have contributed to the final version of the manuscript and agree with its submission to JMIR.

Conflicts of Interest
OP and JB are unpaid members of the scientific advisory committee for the Smoke Free app. LS has received a research grant and honoraria for a talk and travel expenses from manufacturers of smoking cessation medications (Pfizer and Johnson & Johnson). JB and AH have received unrestricted research funding from Pfizer to study smoking cessation.

Multimedia Appendix 1
Sensitivity analyses.

References


Abbreviations
AIC: Akaike information criterion
AR: autoregressive
BF: Bayes factor
COVID-19: coronavirus disease
e-cigarettes: electronic cigarettes
GAMM: generalized additive mixed model
**IRR:** incidence rate ratio

**MA:** moving average

**SA:** sensitivity analysis

**SARS-CoV-2:** severe acute respiratory syndrome coronavirus 2
Developing Empirical Decision Points to Improve the Timing of Adaptive Digital Health Physical Activity Interventions in Youth: Survival Analysis

Adrian Ortega¹, MA; Christopher C Cushing¹,², PhD
¹Clinical Child Psychology Program, University of Kansas, Lawrence, KS, United States
²Schiefelbusch Life Span Institute, University of Kansas, Lawrence, KS, United States

Abstract

Background: Current digital health interventions primarily use interventionist-defined rules to guide the timing of intervention delivery. As new temporally dense data sets become available, it is possible to make decisions about the intervention timing empirically.

Objective: This study aimed to explore the timing of physical activity among youth to inform decision points (eg, timing of support) for future digital physical activity interventions.

Methods: This study comprised 113 adolescents aged between 13 and 18 years (mean age 14.64, SD 1.48 years) who wore an accelerometer for 20 days. Multilevel survival analyses were used to estimate the most likely time of day (via odds ratios and hazard probabilities) when adolescents accumulated their average physical activity. The interacting effects of physical activity timing and moderating variables were calculated by entering predictors, such as gender, sports participation, and school day, into the model as main effects and tested for interactions with the time of day to determine conditional main effects of these predictors.

Results: On average, the likelihood that a participant would accumulate a typical amount of moderate-to-vigorous physical activity increased and peaked between 6 PM and 8 PM before decreasing sharply after 9 PM. Hazard and survival probabilities suggest that optimal decision points for digital physical activity programs could occur between 5 PM and 8 PM.

Conclusions: Overall, the findings of this study support the idea that the timing of physical activity can be empirically identified and that these markers may be useful as intervention triggers.

doi:10.2196/17450

KEYWORDS
telemedicine; exercise; physical activity; adolescent

Introduction

Background

Less than 25% of the youth attain the recommended amounts (60 min per day) of moderate-to-vigorous physical activity (MVPA) [1]. Despite the importance of physical activity for health [2-4], it is well documented that rates of daily MVPA decrease from childhood to adolescence about 40 min each year between ages 9 and 15 years [5]. Accordingly, adolescence is a stage when the youth are likely to exhibit declines in physical activity [6], and therefore, empirical research investigating the patterns of physical activity in adolescence may have value for future interventions to forestall this decline.

The daily activity of youths is in four levels: sleep, sedentary activity, light activity, and MVPA [7,8]. The time spent in each of these levels is strongly linked, such that reducing or increasing the time spent in one of these levels is inversely related to how much time is spent in other levels [9,10]. As the
youth choose to pursue sedentary activities (eg, watching television), the time and opportunity available for exercise decrease. In other words, there is a finite number of minutes in the day, and every passing minute of physical inactivity results in a loss of opportunity for physical activity.

There is growing interest among behavioral scientists to translate efficacious behavioral interventions for improving physical activity into digital modalities [11,12]. Although mobile health interventions for improving health behavior are efficacious, these interventions primarily use technologies to send reminder messages to users [13]. Moreover, high nonusage attrition and declined user engagement throughout the intervention are concerns in the digital health literature [14,15]. Greater intervention precision, such as timely support, may improve digital intervention effectiveness and engagement [16].

Just-in-time adaptive interventions (JITAIs) capitalize on advanced digital technologies and computer automation to tailor digital interventions to the user’s needs and deliver these interventions at opportune moments [16]. A critical component of JITAIs is the decision point or timing of support [16]. JITAIs seek to intervene at critical windows of opportunity for each user to maximize effectiveness and engagement while also minimizing waste or participant burden; however, consistent methods of developing decision points have not been established [16]. In the absence of a clear theory to guide when JITAIs should intervene, it is valuable to empirically identify critical windows of opportunity that can be adjusted based on users’ needs rather than at moments made a priori by investigators as intervention decision points [17]. However, very few digital physical activity interventions have adjusted the timing of intervention delivery based on observed patterns of physical activity [17]. A recent meta-analysis [18] revealed that most physical activity JITAIs modified decision points based on rational decisions, such as interventionist-defined [19-21] or user-defined times [22]. Physical activity JITAIs dynamically tailored their interventions at the individual level based on context-aware sensing and machine learning algorithms [23,24], how and the extent to which the decision points were modified beyond content adaptations, and its unique effect on changes in physical activity remains unclear. In other words, despite these studies, a research gap still remains regarding when adolescents allocate time to physical activity and how we can best leverage these times to improve exercise attainment digitally.

An investigation of the temporal windows of users’ previous exercise patterns can help pinpoint subsequent decision points for intervention. It may prove beneficial to investigate the timing of activity in the youth to identify critical windows of opportunity for when users allocate time to engage in physical activity. Thus, digital interventions can optimize intervention delivery and maximize engagement in physical activity when it is likely to occur for users. By delivering support at the empirically observed moments and at random, interventionist-defined users, or user-selected decision points, may be more likely to exhibit health-promoting behaviors, such as physical activity.

Study Aims and Hypotheses

The primary objective of this study (aim 1) was to explore the time of day when an adolescent accumulated their average physical activity. This aim explored the most likely hour by which an adolescent would have accumulated their average physical activity, given that it had not occurred already. As adolescents are unlikely to meet the recommended 60 min of MVPA each day, this study examined the likelihood that each participant met their average unique MVPA. Aim 1 was exploratory, with no a priori hypothesis postulated.

The secondary objective (aim 2) was to explore day-level and individual-level factors that moderate the likelihood of accumulating physical activity at the time of day found in the primary aim of this study. For this study, the determinants selected for moderation analyses included empirically derived variables previously shown to be correlated with physical activity, such as gender, BMI, sports participation, and school day [25-30]. The primary focus of aim 2 was to determine the magnitude of group differences in the odds of meeting their average MVPA. Gender and BMI are related to exercise attainment and could influence when adolescents meet their average MVPA. For example, because males and individuals with lower BMI engage in more physical activity, they may be more likely to meet their average MVPA compared with females or individuals with a higher BMI. In addition, participation in organized sports or days in which a person attends school might also permit or constrain opportunities to exercise, thereby affecting timing as well. It is hypothesized that youths who participate in sports will obtain higher odds ratios (ORs) of meeting their average MVPA compared with youths who do not participate in sports and that youths will obtain higher ORs of meeting their average MVPA on days in which they attended school than on days in which they did not.

Finally, the third aim (aim 3) was to generate decision points to improve the timing of digital physical activity interventions, given the results from aims 1 and 2. Sets of decision points were inferred from hazard and survival probabilities (see Statistical Analyses section). Decision points were generated for moderators only if the following two conditions were satisfied: (1) if a moderator was found to be significantly different between groups of individuals and (2) if the differences between these groups were as hypothesized in aim 2. These criteria were set forth so that decision points would only be made for subgroups in which there were meaningful differences in the timing of physical activity and to eliminate ineffectual moderators.

Methods

Recruitment

Participants were recruited as part of a study examining the associations between various psychological constructs and physical activity behaviors. Participants learned about the study through flyers posted around the local community. Interested participants were instructed to contact study personnel via phone calls to screen for eligibility. Participants enrolled in the study were aged between 13 and 18 years and lived at home with their caregiver or caregivers. Participants were excluded if they had
any significant physical maladies that would limit physical activity, visual impairments, or an inability to read at their corresponding grade level. These exclusion criteria were in place to ensure a valid assessment of physical activity. A total of 121 adolescents were recruited to participate.

**Procedures**

At the baseline visit, participants reviewed the study information and an institutional review board–approved informed consent form with the research staff and gave consent to participate. For participants aged <18 years, their parents provided informed consent, and adolescents provided informed assent. The participants then completed a demographics questionnaire, a planned activity calendar, and were oriented to the accelerometer. Participants were instructed to wear the accelerometer on their nondominant wrist for the entire 20-day study period [31]. Finally, the research staff measured each participant’s height and weight. Adolescents then wore the accelerometer for 20 days. At the end of the 20 days, participants returned for a laboratory exit visit to return the accelerometers. When all procedures were completed, participants received up to US $40 financial compensation for both laboratory visits and completing the surveys (US $25) and for wearing the accelerometer for at least 18 of the required 20 days (US $15).

Data were collected year-round, beginning in June 2015 and ending in December 2017, across multiple seasons, and when the youth were in and out of school (ie, weekends and weekdays as well as during summer/winter breaks).

**Ethics Approval and Consent to Participate**

In accordance with the revised Common Rule, this study was approved by the sponsoring institution’s human subjects’ protection committee, and all participants provided informed consent and assent.

**Sample Characteristics**

Participants were excluded from the current analyses if they had less than 4 days of valid accelerometer wear (eg, ≥8 hours of accelerometer wear time per day, n=8). Participants in the current analyses (N=113) were aged between 13 and 18 years (mean 14.64, SD 1.48 years). The sample was 37.2% men (32/113) and 62.8% women (71/113). In terms of caregiver demographics, 66.4% (75/113) of parents were married; 22.1% (25/113) were divorced, separated, or widowed; and 11.5% (13/113) were never married. Moreover, 61.1% (69/113) of mothers and 63.7% (72/113) of fathers attained a college education or higher. The majority of the sample (61.1%, 69/113) reported an approximate family income greater than US $60,000. The sample was 78.8% (89/113) Caucasian, 8.0% (9/113) Latino/Latina, 4.4% (5/113) African American, 1.8% (2/113) multiracial, 0.9% (1/113) Native American, and 2.7% (3/113) other.

**Measures and Measurement**

**Demographics**

Adolescents self-reported basic demographic information, including date of birth, age, sex, race, level(s) of parental education, and approximate family income.

**Physical Activity**

The ActiGraph wGT3X-BT accelerometer (ActiGraph LLC) objectively measured participants’ MVPA throughout the duration of the 20-day study period. The accelerometers were programmed to sample movements at a rate of 30 Hz on three different axes, as in previous studies [32] measuring physical activity in the youth. Irrelevant activity periods such as sleep periods and nonwear periods were identified using the Sadeh algorithm and the Troiano algorithm, respectively, and thus removed from the daily activity counts [33,34]. Only days when participants wore the accelerometer for 8 hours or more were included in this study. As in previous studies using wrist-worn accelerometers in youths, the Chandler algorithm was used to identify total minutes of MVPA per day [32,35]. Given previous research demonstrating participant reactivity to accelerometer measurement during the first days of the study, research personnel removed the first 3 days of accelerometer activity data for each participant because of potential activity-based reactivity [36]. After removing these days, the mean of each participant’s unique MVPA was created by calculating the mean of their daily MVPA attainment.

**BMI**

BMI was calculated using the Center for Disease Control and Prevention (CDC) growth charts [37]. Participants’ height in centimeters, weight in kilograms, sex, and age in months at the time of study initiation were used to compute BMI. BMI was calculated using the SAS program for the CDC growth charts found on the CDC website [38].

**Sports Participation**

During the baseline session, participants were asked if they had any planned physical activity, such as involvement in organized sports, for the next 20 days. Similar to previous studies, participants were dichotomously categorized into 2 groups: involved in organized sports and uninvolved in organized sports based on their self-reported activity involvement [25,39-41].

**School Day**

Research personnel classified daily activity patterns into categorical variables: school day and nonschool day to better capture the variability among physical activity patterns for days when youths are in school or not. As in previous retrospective studies investigating the magnitude of physical activity and sedentary time differences when youths are in and out of school, research personnel used the school calendar to classify school days vs nonschool days [42].

**Statistical Analyses**

To address the aims of this study, multilevel survival analyses using logistic regression were conducted to examine the hour of day when adolescents accumulated their average physical activity. Predictors were entered into the model as main effects and subsequently tested as interactions with the time of day to determine the conditional main effects of study moderators on the time of day when adolescents accumulated their average physical activity. For the purposes of this project, the time of day was analyzed as a discrete time variable. Using a special case of logistic regression, a hazard function, or the probability
of the event occurring before the time (hour of day, in this case), conditional on no earlier occurrence, was estimated [43].

To estimate hazard functions, multiple smoothing procedures using polynomial functions of time were tested [43]. For each survival analysis, polynomials of time were entered into the model as linear, quadratic, and cubic predictors of the event. Each survival analysis was estimated using maximum likelihood estimation based on Laplace approximation in SAS PROC GLIMMIX. Nested model comparisons using the chi-square difference of the estimated $-2 \times \log$ likelihood ratio tests for each of these models were evaluated to determine the best fitting model for the polynomial effect of time. Predictors as well as interactions between predictors and time of day were then entered into the hazard model. Each of these hazard models was estimated separately. In the event of a significant interaction of a predictor and time of day, the effect of these interactions was determined by comparing the OR of the estimates by summing all parameter estimates multiplied by their respective variables and then using the inverse link function (ie, $e^{\log (y=1)}$) to translate estimated logits into ORs. To compare the estimates, ORs were calculated by inserting meaningful values into the explanatory predictors in the regression equations. To address aims 1 and 2 OR estimates were used to determine the most likely hour of average MVPA accumulation and group differences in the timing of MVPA obtainment using 8 AM as the reference hour. To investigate whether there were significant differences in the magnitude of the ORs for the time of average MVPA accumulation between groups at $P$ values <.05 level of significance, the 95% CIs around the estimated ORs were compared. OR estimates that did not overlap based on their given 95% confidence bands were considered statistically significant at the $P$ value <.05 level [44]. For each moderator, statistical differences between groups at every hour of the day were examined.

If the ORs were statistically significant for a window of time in the hypothesized direction, decision points using hazard and survival probabilities were developed, as stated in aim 3. The hazard and survival probabilities were created by translating the estimated logits into probabilities. Survival probabilities represent the cumulative risk that an individual would not have met the event at a certain time. To estimate the survival probability at each hour, we multiplied the complement of the hazard probabilities (ie, $1 \times$ probability) for that hour and all previous hours [45].

**Results**

### Preliminary Analyses

On average, participants accumulated 30.91 min (SD 30.94; range 0-98.16) of MVPA per day and, therefore, did not meet the recommended guidelines. Participants wore the accelerometer for an average of 17.60 (SD 4.5; range 1-25) valid wear days.

### Aim 1

After inserting sequential polynomials of time, nested model comparisons indicated that a cubic function of time was the best fitting model. On average, the likelihood that a participant would accumulate their own average MVPA increased and peaked between 6 PM and 8 PM (OR 13.19-13.02) before decreasing sharply after 9 PM (Table 1 and Figure 1). No participants met their average MVPA before 8 AM or after 11 PM. For this reason, tables and figures displaying results for aims 1 to 3 reflect the ORs, hazard probabilities, and survival probabilities that an individual met their average MVPA by that hour, conditional on no earlier occurrence, beginning between the hours before 8 AM and terminating before 11 PM.
Table 1. Odds ratio estimates of obtaining average moderate-to-vigorous physical activity before hour of day.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Frequency^a</th>
<th>Odds ratio estimates (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 8 AM (reference)</td>
<td>0</td>
<td>N/A^b</td>
</tr>
<tr>
<td>Before 9 AM</td>
<td>5</td>
<td>1.24 (1.31-1.36)</td>
</tr>
<tr>
<td>Before 10 AM</td>
<td>21</td>
<td>1.60 (1.34-1.91)</td>
</tr>
<tr>
<td>Before 11 AM</td>
<td>18</td>
<td>2.13 (1.67-2.72)</td>
</tr>
<tr>
<td>Before noon</td>
<td>24</td>
<td>2.87 (2.12-3.7)</td>
</tr>
<tr>
<td>Before 1 PM</td>
<td>22</td>
<td>3.90 (2.73-5.56)</td>
</tr>
<tr>
<td>Before 2 PM</td>
<td>24</td>
<td>5.24 (3.53-7.77)</td>
</tr>
<tr>
<td>Before 3 PM</td>
<td>50</td>
<td>6.90 (4.51-10.55)</td>
</tr>
<tr>
<td>Before 4 PM</td>
<td>55</td>
<td>8.79 (5.63-13.72)</td>
</tr>
<tr>
<td>Before 5 PM</td>
<td>85</td>
<td>10.70 (6.78-16.89)</td>
</tr>
<tr>
<td>Before 6 PM</td>
<td>89</td>
<td>12.30 (7.77-19.45)</td>
</tr>
<tr>
<td>Before 7 PM</td>
<td>82</td>
<td>13.19 (8.38-20.73)</td>
</tr>
<tr>
<td>Before 8 PM</td>
<td>91</td>
<td>13.02 (8.38-20.22)</td>
</tr>
<tr>
<td>Before 9 PM</td>
<td>79</td>
<td>11.70 (7.66-17.84)</td>
</tr>
<tr>
<td>Before 10 PM</td>
<td>43</td>
<td>9.45 (6.3-14.12)</td>
</tr>
<tr>
<td>Before 11 PM</td>
<td>32</td>
<td>6.77 (4.6-9.94)</td>
</tr>
</tbody>
</table>

^aFrequency represents the number of instances in which users met their average moderate-to-vigorous physical activity before hour of day.

^bN/A: not applicable.

Figure 1. Odds ratios of obtaining average moderate-to-vigorous physical activity before each hour of day.

Aim 2

Compared with 8 AM, male adolescents had significantly higher odds of obtaining their average MVPA only between the 8 AM and noon window compared with female adolescents (Table 2 and Figure 2).

BMI did not significantly moderate the relationship between time of day and MVPA attainment (Table 3 and Figure 3).

In addition, adolescents involved in organized sports had significantly lower odds of attaining their average MVPA between 8 AM and 10 AM (Table 4 and Figure 4).

Adolescents had significantly higher odds between the 8 AM and noon window, compared with 8 AM, of meeting their average MVPA on nonschool days compared with school days (Table 5 and Figure 5).
Table 2. Moderating effect of sex on odds ratio estimates of time of average moderate-to-vigorous physical activity accumulation.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Odds ratio estimates (95% CI)</th>
<th>Males(^a) (n=42)</th>
<th>Females(^b) (n=71)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 8 AM (reference)</td>
<td>N/A(^c)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Before 9 AM</td>
<td>1.52 (1.29-1.79)(^d)</td>
<td>1.12 (1.1-1.25)</td>
<td></td>
</tr>
<tr>
<td>Before 10 AM</td>
<td>2.30 (1.8-3.34)(^d)</td>
<td>1.33 (1.08-1.64)</td>
<td></td>
</tr>
<tr>
<td>Before 11 AM</td>
<td>3.43 (2.42-5.78)(^d)</td>
<td>1.66 (1.24-2.23)</td>
<td></td>
</tr>
<tr>
<td>Before noon</td>
<td>5.02 (3.21-9.54)(^d)</td>
<td>2.16 (1.49-3.11)</td>
<td></td>
</tr>
<tr>
<td>Before 1 PM</td>
<td>7.15 (4.2-14.95)</td>
<td>2.86 (1.85-4.38)</td>
<td></td>
</tr>
<tr>
<td>Before 2 PM</td>
<td>9.83 (4.87-16.66)</td>
<td>3.81 (2.35-6.11)</td>
<td></td>
</tr>
<tr>
<td>Before 3 PM</td>
<td>12.94 (6.07-27.23)</td>
<td>5.04 (3.8-36)</td>
<td></td>
</tr>
<tr>
<td>Before 4 PM</td>
<td>16.20 (7.31-35.35)</td>
<td>6.52 (3.76-11.04)</td>
<td></td>
</tr>
<tr>
<td>Before 5 PM</td>
<td>19.12 (8.43-42.52)</td>
<td>8.11 (4.61-13.88)</td>
<td></td>
</tr>
<tr>
<td>Before 6 PM</td>
<td>21.13 (9.26-47.12)</td>
<td>9.58 (5.41-16.38)</td>
<td></td>
</tr>
<tr>
<td>Before 7 PM</td>
<td>21.69 (9.57-47.79)</td>
<td>10.58 (6-19.91)</td>
<td></td>
</tr>
<tr>
<td>Before 8 PM</td>
<td>20.52 (9.24-44.12)</td>
<td>10.77 (6.18-17)</td>
<td></td>
</tr>
<tr>
<td>Before 9 PM</td>
<td>17.76 (8.23-36.87)</td>
<td>9.96 (5.81-16.15)</td>
<td></td>
</tr>
<tr>
<td>Before 10 PM</td>
<td>13.96 (6.7-27.81)</td>
<td>8.24 (4.89-13.01)</td>
<td></td>
</tr>
<tr>
<td>Before 11 PM</td>
<td>9.88 (4.9-18.91)</td>
<td>6.02 (3.62-9.27)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Average moderate-to-vigorous physical activity for men=31.48 min.
\(^b\)Average moderate-to-vigorous physical activity for women=30.66 min.
\(^c\)N/A: not applicable.
\(^d\)Significant differences in odds ratios between groups based on nonoverlapping CIs.

Figure 2. Sex differences in odds ratios of obtaining average moderate-to-vigorous physical activity before each hour of day.
### Table 3. Moderating effect of BMI on odds ratio estimates of time of average moderate-to-vigorous physical activity accumulation.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Normal weight status(^a) (n=81)</th>
<th>Overweight status(^b) (n=13)</th>
<th>Obese status(^c) (n=19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 8 AM (reference)</td>
<td>N/A(^d)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Before 9 AM</td>
<td>1.28 (1.16-1.39)</td>
<td>1.25 (1.14-1.36)</td>
<td>1.24 (1.24-1.36)</td>
</tr>
<tr>
<td>Before 10 AM</td>
<td>1.68 (1.4-1.97)</td>
<td>1.61 (1.36-1.9)</td>
<td>1.59 (1.34-1.89)</td>
</tr>
<tr>
<td>Before 11 AM</td>
<td>2.26 (1.76-2.83)</td>
<td>2.14 (1.68-2.7)</td>
<td>2.10 (1.64-2.68)</td>
</tr>
<tr>
<td>Before noon</td>
<td>3.07 (2.25-4.07)</td>
<td>2.86 (2.12-3.83)</td>
<td>2.8 (2.06-3.8)</td>
</tr>
<tr>
<td>Before 1 PM</td>
<td>4.16 (2.91-5.81)</td>
<td>3.83 (2.7-5.4)</td>
<td>3.74 (2.61-5.33)</td>
</tr>
<tr>
<td>Before 2 PM</td>
<td>5.57 (3.76-8.12)</td>
<td>5.06 (3.42-7.44)</td>
<td>4.93 (3.3-7.35)</td>
</tr>
<tr>
<td>Before 3 PM</td>
<td>7.29 (4.79-10.99)</td>
<td>6.54 (4.29-9.92)</td>
<td>6.34 (4.09-9.8)</td>
</tr>
<tr>
<td>Before 4 PM</td>
<td>9.22 (5.95-14.24)</td>
<td>8.15 (5.23-12.67)</td>
<td>7.87 (4.95-12.5)</td>
</tr>
<tr>
<td>Before 5 PM</td>
<td>11.14 (7.17-17.48)</td>
<td>9.68 (6.13-15.29)</td>
<td>9.30 (5.73-15.09)</td>
</tr>
<tr>
<td>Before 6 PM</td>
<td>12.73 (8.11-20.11)</td>
<td>10.83 (6.82-17.27)</td>
<td>10.35 (6.3-17.03)</td>
</tr>
<tr>
<td>Before 7 PM</td>
<td>13.58 (8.7-21.44)</td>
<td>11.29 (7.11-18.04)</td>
<td>10.71 (6.48-17.79)</td>
</tr>
<tr>
<td>Before 8 PM</td>
<td>13.40 (8.66-20.99)</td>
<td>10.82 (6.85-17.25)</td>
<td>10.18 (6.14-17.01)</td>
</tr>
<tr>
<td>Before 9 PM</td>
<td>12.09 (7.91-18.67)</td>
<td>9.43 (6.01-14.95)</td>
<td>8.79 (5.29-14.74)</td>
</tr>
<tr>
<td>Before 10 PM</td>
<td>9.86 (6.53-14.96)</td>
<td>7.39 (4.74-11.64)</td>
<td>6.80 (4.09-11.47)</td>
</tr>
<tr>
<td>Before 11 PM</td>
<td>7.19 (4.79-10.72)</td>
<td>5.13 (3.31-8.07)</td>
<td>4.66 (2.79-7.94)</td>
</tr>
</tbody>
</table>

\(^a\)Mean moderate-to-vigorous physical activity (MVPA) for normal weight status=31.23 min.

\(^b\)Mean MVPA for overweight status=27.33 min.

\(^c\)Mean MVPA for obese status=32.32.

\(^d\)N/A: not applicable.
Figure 3. BMI differences in odds ratios of obtaining average moderate-to-vigorous physical activity before each hour of day.
Table 4. Moderating effect of involvement in organized sports (sports participation) on odds ratio estimates of time of average moderate-to-vigorous physical activity accumulation.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Odds ratio estimates (95% CI)</th>
<th>No sports participation&lt;sup&gt;a&lt;/sup&gt; (n=68)</th>
<th>Sports participation&lt;sup&gt;b&lt;/sup&gt; (n=45)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td>Before 8 AM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 9 AM</td>
<td></td>
<td>1.41 (1.26-1.58)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.95 (0.81-1.12)</td>
</tr>
<tr>
<td>Before 10 AM</td>
<td></td>
<td>1.95 (1.62-2.48)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.05 (0.78-1.43)</td>
</tr>
<tr>
<td>Before 11 AM</td>
<td></td>
<td>2.65 (1.96-3.57)</td>
<td>1.32 (0.86-2.02)</td>
</tr>
<tr>
<td>Before noon</td>
<td></td>
<td>3.51 (2.42-5.1)</td>
<td>1.83 (1.16-3.41)</td>
</tr>
<tr>
<td>Before 1 PM</td>
<td></td>
<td>4.54 (2.94-7)</td>
<td>2.68 (1.44-4.96)</td>
</tr>
<tr>
<td>Before 2 PM</td>
<td></td>
<td>5.71 (3.53-9.25)</td>
<td>4.06 (2.2-8.9)</td>
</tr>
<tr>
<td>Before 3 PM</td>
<td></td>
<td>6.96 (4.15-11.7)</td>
<td>6.14 (2.9-12.81)</td>
</tr>
<tr>
<td>Before 4 PM</td>
<td></td>
<td>8.22 (4.78-14.15)</td>
<td>9.01 (4.41-21.43)</td>
</tr>
<tr>
<td>Before 5 PM</td>
<td></td>
<td>9.36 (5.37-16.33)</td>
<td>12.44 (5.24-27.36)</td>
</tr>
<tr>
<td>Before 6 PM</td>
<td></td>
<td>10.26 (5.88-17.95)</td>
<td>15.65 (7.34-37.64)</td>
</tr>
<tr>
<td>Before 7 PM</td>
<td></td>
<td>10.79 (6.23-18.75)</td>
<td>17.40 (7.7-37.94)</td>
</tr>
<tr>
<td>Before 8 PM</td>
<td></td>
<td>10.88 (6.38-18.6)</td>
<td>16.57 (7.92-38.37)</td>
</tr>
<tr>
<td>Before 9 PM</td>
<td></td>
<td>10.47 (6.28-17.52)</td>
<td>13.11 (6.05-27.03)</td>
</tr>
<tr>
<td>Before 10 PM</td>
<td></td>
<td>9.60 (5.91-15.68)</td>
<td>8.35 (4.28-18.27)</td>
</tr>
<tr>
<td>Before 11 PM</td>
<td></td>
<td>8.36 (5.47-13.96)</td>
<td>4.16 (2.02-7.98)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mean moderate-to-vigorous physical activity (MVPA) for nonsports participators=24.45 min.

<sup>b</sup>Mean MVPA for sports participators=40.81 minutes.

<sup>c</sup>N/A: not applicable.

<sup>d</sup>Significant differences in odds ratios between groups based on nonoverlapping CIs.
Figure 4. Involvement in organized sports differences in odds ratios of obtaining average moderate-to-vigorous physical activity before each hour of day.
Table 5. Moderating effect of school day on odds ratio estimates of time of average moderate-to-vigorous physical activity accumulation.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Odds ratio estimates (95% CI)</th>
<th>School day&lt;sup&gt;b&lt;/sup&gt; (862 days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonschool day&lt;sup&gt;a&lt;/sup&gt; (1036 days)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Before 8 AM (reference)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N/A</td>
</tr>
<tr>
<td>Before 9 AM</td>
<td>1.55 (1.63-1.77)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.99 (0.87-1.12)</td>
</tr>
<tr>
<td>Before 10 AM</td>
<td>2.34 (1.82-3)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.1 (0.86-1.4)</td>
</tr>
<tr>
<td>Before 11 AM</td>
<td>3.41 (2.4-4.83)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.34 (0.95-1.88)</td>
</tr>
<tr>
<td>Before noon</td>
<td>4.79 (3.08-7.39)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.76 (1.15-2.68)</td>
</tr>
<tr>
<td>Before 1 PM</td>
<td>6.48 (3.87-10.73)</td>
<td>2.43 (1.47-3.95)</td>
</tr>
<tr>
<td>Before 2 PM</td>
<td>8.43 (4.73-14.75)</td>
<td>3.44 (1.97-5.89)</td>
</tr>
<tr>
<td>Before 3 PM</td>
<td>10.50 (5.63-19.17)</td>
<td>4.88 (2.67-8.67)</td>
</tr>
<tr>
<td>Before 4 PM</td>
<td>12.52 (6.5-23.48)</td>
<td>6.77 (3.59-12.31)</td>
</tr>
<tr>
<td>Before 5 PM</td>
<td>14.25 (7.25-27.09)</td>
<td>9.00 (4.67-16.51)</td>
</tr>
<tr>
<td>Before 6 PM</td>
<td>15.44 (7.8-29.37)</td>
<td>11.16 (5.75-20.43)</td>
</tr>
<tr>
<td>Before 7 PM</td>
<td>15.91 (8.06-29.9)</td>
<td>12.65 (6.53-22.82)</td>
</tr>
<tr>
<td>Before 8 PM</td>
<td>15.54 (7.99-28.56)</td>
<td>12.78 (6.66-22.52)</td>
</tr>
<tr>
<td>Before 9 PM</td>
<td>14.37 (7.54-25.59)</td>
<td>11.26 (5.95-19.25)</td>
</tr>
<tr>
<td>Before 10 PM</td>
<td>12.54 (6.76-21.54)</td>
<td>8.44 (5.53-13.99)</td>
</tr>
<tr>
<td>Before 11 PM</td>
<td>10.32 (5.7-17.08)</td>
<td>5.27 (2.85-8.5)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Mean moderate-to-vigorous physical activity (MVPA) on nonschool days=28.49 min.
<sup>b</sup>Mean MVPA on school days=33.82 min.
<sup>c</sup>N/A: not applicable.
<sup>d</sup>Significant differences in ORs between groups based on nonoverlapping CIs.
Aim 3

On average, hazard probabilities indicated that adolescents were most likely to instantaneously meet their average MVPA between 5 PM and 8 PM (Table 6). OR estimates and hazard probabilities indicate slightly different times owing to their mathematical computation. The most likely times offered by the OR estimates and hazard probabilities generally overlap and are conceptually congruent. Survival probabilities demonstrate that after 8 PM, adolescents had a 73% chance of not meeting their own MVPA average (Table 6 and Figure 6). The sharp decline (~12%) in survival probabilities between 5 PM and 8 PM indicates that adolescents’ risk of not meeting their average MVPA drastically reduced during this time compared with the windows of time before and after this period. Intervention decision points should therefore be prioritized during this period (5 PM to 8 PM).

On the basis of the criteria set in place for generating decision points for moderators, sex was the only moderator that upheld both criteria. Although men were significantly more likely to meet their average MVPA during the 8 AM to noon time, there are essentially no differences in the hazard and survival probabilities of MVPA accumulation during this window. Nonetheless, sex differences in the OR estimates at these times indicate that the timing of activity may differ across sexes. Overall, male adolescents might benefit from additional digital support for exercise during the morning hours compared with females.
Table 6. Hazard and survival probabilities of time of average moderate-to-vigorous physical activity accumulation.

<table>
<thead>
<tr>
<th>Hour</th>
<th>Hazard probability</th>
<th>Survival probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 8 AM</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Before 9 AM</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Before 10 AM</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Before 11 AM</td>
<td>0.01</td>
<td>0.98</td>
</tr>
<tr>
<td>Before noon</td>
<td>0.01</td>
<td>0.97</td>
</tr>
<tr>
<td>Before 1 PM</td>
<td>0.02</td>
<td>0.96</td>
</tr>
<tr>
<td>Before 2 PM</td>
<td>0.03</td>
<td>0.94</td>
</tr>
<tr>
<td>Before 3 PM</td>
<td>0.03</td>
<td>0.91</td>
</tr>
<tr>
<td>Before 4 PM</td>
<td>0.04</td>
<td>0.88</td>
</tr>
<tr>
<td>Before 5 PM</td>
<td>0.05</td>
<td>0.85</td>
</tr>
<tr>
<td>Before 6 PM</td>
<td>0.05</td>
<td>0.81</td>
</tr>
<tr>
<td>Before 7 PM</td>
<td>0.05</td>
<td>0.77</td>
</tr>
<tr>
<td>Before 8 PM</td>
<td>0.04</td>
<td>0.73</td>
</tr>
<tr>
<td>Before 9 PM</td>
<td>0.04</td>
<td>0.70</td>
</tr>
<tr>
<td>Before 10 PM</td>
<td>0.03</td>
<td>0.68</td>
</tr>
<tr>
<td>Before 11 PM</td>
<td>0.00</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Figure 6. Survival probabilities of obtaining average moderate-to-vigorous physical activity before each hour of day.
Discussion

Principal Findings

The purpose of this study was to explore the timing of exercise for adolescents, identify correlates of physical activity that moderate the timing of exercise, and generate decision points for digital physical activity interventions. No other studies were found that explored the time of day when adolescents meet their typical levels of MVPA or the time of day when adolescents generally exercise. The pattern of ORs, hazard probabilities, and survival probabilities indicating typical MVPA attainment in the late afternoon and early evening (5 PM to 8 PM) coincides with adolescents’ daily schedule. Generally, adolescents’ ability to exercise is constrained by their school attendance, designating early-to-mid afternoon as the earliest convenience for adolescents to meaningfully accrue MVPA [46]. Furthermore, the decline in ORs, as well as hazard probabilities after 8 PM can be explained by adolescents’ needs to attend to other important routines, such as eating dinner, completing homework, and preparing for the next day [47]. Therefore, adolescents may be less likely to engage with digital health intervention options outside of this window of time, and digital support to encourage exercise during these times could be wasteful. This persistent inopportune support is likely to lead to intervention failure and decreased user engagement, a continual challenge in the digital health literature [14-16,48]. These findings suggest that just-in-time support during this window could be most helpful for adolescents and lead to positive engagement with digital support for exercise.

These findings show that males have significantly higher odds of obtaining their average MVPA in the morning compared with females. These findings highlight that daily exercise patterns may vary by sex. Furthermore, these results suggest that male adolescents may benefit from exercise prompts in the morning and in the afternoon, whereas female adolescents might only benefit from exercise prompts later in the day. In other words, although it could be wasteful to prompt female adolescents to exercise in the morning, supplementary digital support for exercise in the morning (eg, 8 AM to noon) could be helpful for male adolescents. Although sending digital support for exercise in the morning to males could be conceived as wasteful, given that they are in school, research demonstrates that male adolescents obtain more exercise than female adolescents in school, including when leisurely on school grounds and also when at recess and gym class [49]. Therefore, digital support for male teenagers during this time may prove beneficial.

There were no significant differences in the timing of the attainment of typical MVPA across groups of different weight statuses. Although previous research has indicated that youths with overweight and obesity exercise less, this study did not find substantial differences in MVPA levels across weight statuses (Table 3) [25,28]. The absence of MVPA differences across groups of varying weight statuses probably contributed to the lack of timing differences across these groups. In this case, BMI may not affect the actual timing of exercise. Surprisingly, nonsports participants had significantly higher odds of obtaining their typical MVPA in the morning compared with sports participants. Consistent with previous research, sports participants in this study displayed a higher mean MVPA (Table 4) [50]. Therefore, it should take longer for sports participants to accumulate their typical levels of MVPA compared with nonsports participants. This relationship between higher MVPA averages and later timing of typical MVPA attainment may explain why these results contradict the hypothesis. In addition, this finding reflects adolescents’ typical sports practice and game schedule such that sports participants may be engaging in MVPA later in the day after school practices or evening games.

There were higher odds of MVPA attainment in the morning on nonschool days than on school days, which may be indicative of the lack of constraints that prevent exercise on school days, especially considering that there were similar levels of mean MVPA across school vs nonschool days (Table 5). In other words, there may be more opportunities for adolescents to exercise in the morning on nonschool days, which might modify the timing of support on those days [30].

On the basis of this study, decision points for JITAIs promoting exercise could occur between the 5 PM and 8 PM time frame and between 8 AM and noon for male teenagers, as indicated by ORs, hazard probabilities, and survival probabilities. This period appears to overlap with adolescent exercise patterns and could serve as an optimal starting place for novice exercisers to accrue MVPA. In addition, on days when youths have met their typical MVPA by this window, this period could serve as an opportunity to make exercise gains. It should be added that survival probabilities indicated that youths are still 73% unlikely to obtain their typical MVPA after 8 PM. There were more days when adolescents did not meet their average than the days they did. This finding indicates that encouraging adolescents to consistently meet their own MVPA average would constitute a meaningful shift in MVPA attainment and a consequent increase in their typical MVPA levels, which could arrest the decline in MVPA observed during adolescence.

Limitations and Future Directions

To determine when users engaged in exercise, this study investigated the time of day when adolescents met their mean MVPA. Some users probably do not accrue MVPA in extensive, continuous bouts of time; rather, they likely obtain MVPA over intermittent spans of time throughout the day [46]. For instance, a person may gain some MVPA walking to school in the morning, in gym class, and after school in sports practice. Therefore, users may engage in exercise multiple times throughout the day, and this study’s conceptualization of physical activity timing does not capture this pattern. Furthermore, the timing of obtaining a typical MVPA is likely to be earlier for individuals who typically accrue little-to-no MVPA compared with those who typically accrue more MVPA. Given that these results are averages across individuals, it is also important to acknowledge the potential inability to generalize this average physical activity timing profile to other samples and populations. In addition, interventions based on these results need to consider individual variability in physical activity timing based on their unique activity patterns. In other
words, the decision points to improve exercise might differ between sedentary adolescents and active adolescents. For example, because sedentary adolescents might need more prompts to exercise or might have already obtained their typical levels of MVPA earlier in the day, decision points for a sedentary adolescent could occur more frequently, such as after prolonged bouts of sedentary time, whereas a more active adolescent might benefit from decision points at the times when they are usually physically active.

This study included participants recruited from different seasons of the year (eg, winter and summer), and although season may affect the timing of MVPA, by distributing data collection throughout the year, our findings are more generalizable than if they were taken from a single season. In this study, weekend days during the school year and weekdays during the summer were both classified as nonschool days, given that the lack of school during both these types of days could similarly affect the timing of activity. However, this study did not evaluate the seasonal effect of physical activity timing, which is a limitation of the study. In addition, work hours or other contexts that would constrain an adolescent’s ability to engage in MVPA were not assessed in the study. Future research should consider how these contexts would suppress one’s ability to exercise and ultimately affect their decision points.

Each of the moderators in aim 2 was analyzed independently of the others. In reality, these variables are not mutually exclusive and interact, such that there could have been unlimited interactions between these variables that moderated the most likely time of exercise. Future research should consider a more nuanced examination of how these moderators in tandem influence the timing of exercise to better optimize decision points for physical activity JITAIs across multiple contexts. Another limitation of the study is that it did not seek to determine which moderator of timing would be the most useful for adjusting decision points. Therefore, future research should investigate the experimental effects of tailoring decision points via different situational and contextual factors on improvements in physical activity.

Furthermore, it is possible that other variables might moderate the timing of exercise, including one’s built environment characteristics (eg, neighborhood walkability and access to recreational activities; [51-53]). For example, youths who actively transport (eg, biking and walking) themselves to and from school might engage in more MVPA during these windows. Moreover, the moderators evaluated in this study are typically time-varying factors, influence the timing of exercise. Such an examination of how these moderators in tandem interact, such that there could have been unlimited interactions between these variables that moderated the most likely time of exercise. Future research should consider how these contexts would suppress one’s ability to exercise and ultimately affect their decision points.

Developing decision points for JITAIs by investigating the timing of exercise is a direct answer to calls in the research literature to model and incorporate microtemporal dynamics of health determinants or the study of behavioral phenomena in small timescales, within health behavior science [16,32,55]. Expansion of health behavior research to the microtimescale may elucidate the temporal specificity of health behaviors such as physical activity [16,55]. In addition, because the microtemporal study of physical activity can be continuously monitored and passively detected, JITAIs can therefore optimize and adapt interventions for each individual user more readily than in-person and static digital interventions. Most importantly, with these temporally dense data sets, researchers can statistically uncover the unique temporal patterns for each participant, given the immense quantity of observations [53,56].

It is conceivable that activity patterns may be highly idiosyncratic depending on the temporal, contextual, and psychosocial processes involved. However, with these idiosyncratic data, automated JITAIs can be tailored to match the temporal specificity of each individual user’s microtemporal physical activity patterns. Future research exploring the timing of exercise and decision points generally should, therefore, consider employing idiosyncratic methods, such as precision medicine approaches, to enhance decision points for each user [57,58].

Ultimately, the notion that improved timing of support for digital physical activity interventions will lead to improvements in physical activity requires experimental testing. This study demonstrates that decision points can be empirically defined and that the timing of exercise may differ among groups of people but did not evaluate the impact of tailoring these decision points on adolescent physical activity. To address this gap, it should be empirically tested if sending digital support at empirically identified moments leads to more exercise compared with digital support at interventionist-defined or user-defined
times. Major areas for future research include investigating the impact of sending digital support at the empirically identified window (5 PM to 8 PM) on physical activity, evaluating adolescent receptivity to digital support during this time as well as generating and tailoring idiographic decision points based on an individual’s unique physical activity patterns.

Acknowledgments
This project was supported by a targeted research grant from the Society of Pediatric Psychology awarded to the second author. The paper processing charges related to the publication of this study were supported by the University of Kansas (KU) One University Open Access Author Fund sponsored jointly by the KU Provost, KU Vice Chancellor for Research and Graduate Studies, and KU Medical Center Vice Chancellor for Research and managed jointly by the Libraries at the Medical Center and KU, Lawrence.

Conflicts of Interest
None declared.

References


https://mhealth.jmir.org/2020/6/e17450


Abbreviations

CDC: Center for Disease Control and Prevention

JITAI: just-in-time adaptive intervention

KU: University of Kansas

https://mhealth.jmir.org/2020/6/e17450

JMIR Mhealth Uhealth 2020 | vol. 8 | iss. 6 | e17450 | p.575

(page number not for citation purposes)
**MVPA**: moderate-to-vigorous physical activity

**OR**: odds ratio

Edited by G Eysenbach; submitted 12.12.19; peer-reviewed by A Psilogios, K Devine, D Saleh; comments to author 26.01.20; revised version received 09.02.20; accepted 26.02.20; published 10.06.20.

Please cite as:
Ortega A, Cushing CC
Developing Empirical Decision Points to Improve the Timing of Adaptive Digital Health Physical Activity Interventions in Youth: Survival Analysis
JMIR Mhealth Uhealth 2020;8(6):e17450
URL: https://mhealth.jmir.org/2020/6/e17450
doi:10.2196/17450
PMID:32519967

©Adrian Ortega, Christopher C Cushing. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 10.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Telemedicine in Chronic Wound Management: Systematic Review And Meta-Analysis

Lihong Chen¹, MD; Lihui Cheng², BSc; Wei Gao¹, MD; Dawei Chen¹, MD; Chun Wang¹, MD; Xingwu Ran¹, MD

¹Diabetic Foot Care Center, Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, Chengdu, China
²Department of Clinical Laboratory, Nuclear Industry 416 Hospital, the 2nd Affiliated Hospital of Chengdu Medical College, Chengdu, China
³School of Clinical Medicine, Southwest Medical University, Luzhou, China

Corresponding Author:
Xingwu Ran, MD
Diabetic Foot Care Center, Department of Endocrinology and Metabolism
West China Hospital
Sichuan University
37 Guo Xue Lane
Chengdu,
China
Phone: 86 18980601305
Email: ranxingwu@163.com

Abstract

Background: Chronic wounds have been a great burden to patients and the health care system. The popularity of the internet and smart devices, such as mobile phones and tablets, has made it possible to adopt telemedicine (TM) to improve the management of chronic wounds. However, studies conducted by different researchers have reported contradictory results on the effect of TM on chronic wound management.

Objective: The aim of this work was to evaluate the efficacy and safety of TM in chronic wound management.

Methods: We systematically searched multiple electronic databases (MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials [CENTRAL]) to identify eligible studies published from inception to June 12, 2019. Inclusion criteria were randomized controlled trials (RCTs) and interventional cohort studies that investigated the use of TM in chronic wound management. RCT and observational data were analyzed separately. A meta-analysis and qualitative analysis were conducted to estimate endpoints.

Results: A total of 6 RCTs and 6 cohort studies including 3913 patients were included. Of these, 4 studies used tablets or mobile phones programmed with apps, such as Skype and specialized interactive systems, whereas the remaining 8 studies used email, telephone, and videoconferencing to facilitate the implementation of TM using a specialized system. Efficacy outcomes in RCTs showed no significant differences in wound healing (hazard ratio [HR] 1.16, 95% CI 0.96-1.39; P=.13), and wound healing around 1 year (risk ratio [RR] 1.05, 95% CI 0.89-1.23; P=.15). Noninferiority criteria of TM were met. A decreased risk of amputation in patients receiving TM was revealed (RR 0.45, 95% CI 0.29-0.71; P=.001). The result of cohort studies showed that TM was more effective than standard care (HR 1.74, 95% CI 1.43-2.12; P<.001), whereas the outcome efficacy RR of wound healing around 1 year (RR 1.21, 95% CI 0.96-1.53; P=.56) and 3 months (RR 1.24, 95% CI 0.47-3.3; P=.67) was not significantly different between TM and standard care. Noninferiority criteria of TM were met for wound healing around 1 year in cohort studies.

Conclusions: Currently available evidence suggests that TM seems to have similar efficacy and safety, and met noninferiority criteria with conventional standard care of chronic wounds. Large-scale, well-designed RCTs are warranted.

(JMIR Mhealth Uhealth 2020;8(6):e15574) doi:10.2196/15574

KEYWORDS

telemedicine; wounds and injuries; wound healing; meta-analysis
Introduction

A chronic wound is defined as a break in the skin that failed to progress through a normal sequence of repair in 4-8 weeks [1]. Venous stasis ulcers, arterial ulcers, neuropathic ulcers, pressure ulcers, diabetic foot ulcers, and ulcers due to malignancy are examples of chronic wounds. This has become a great challenge and burden to patients, health care professionals, and the health care systems. There are over 6 million chronic wounds in the United States, which accounts for an estimated US $25 billion annually in the US health care costs [2]. Thus, there is huge pressure on the health care system to develop cost-effective wound management practices.

Telemedicine (TM) is the use of telecommunication technologies to provide remote clinical services to patients to improve the quality of individual treatments. The concept of TM dates back to the 19th century. It was practiced via telegraph, telephone, and radio before the internet existed [3,4]. With the ascent of the information age, networks and smartphones have shown great potential in providing remote clinical services. It has been widely used in various areas of health care such as heart conditions [5], diabetes mellitus care [6], and management of individuals with chronic obstructive pulmonary disease [7].

Within wound care, TM could support access to expertise in remote areas to improve management of chronic wounds in geographically challenging environments [8,9]. Because of the lack of wound care specialists and the financial pressure on health care agencies, TM could be introduced as an alternative solution to support task shifting of experts from hospitals to underserved populations or rural areas. This could contribute to a reduction in the number of consultations and associated transportation costs [10]. It is also important to realize these goals without impinging on the quality of care. Furthermore, the development of mobile phone apps make it convenient to implement the TM for diabetic foot ulcers [11,12].

Findings from qualitative studies show positive results with several systematic reviews in recent years being published [13-16]. With the convenience and accuracy of communication in the information age, there has been a growing interest in the management of chronic wound via TM. However, the impact of TM on wound healing was inconsistent, with 2 reports indicating positive results [17,18], 1 negative [19], and 3 showing no change [20-22]. In terms of mortality and amputation, 2 RCTs revealed inconsistent results [20,22]. Thus, there is a rational to conduct a systematic review to clarify the effect of TM on chronic wounds.

Our objective was to conduct a systematic review and meta-analysis of randomized and interventional cohort studies that investigated the use of TM follow-up in the community care in collaboration with specialists in wound centers with a comparator of no TM. Only articles that investigated chronic wounds, such as diabetic foot ulcers, stasis ulcer, pressure ulcers, and nonhealing surgical wound, were included. Exclusion criteria included case reports, editorials, letters, animal studies, case–control studies, and self-control studies. Non-English articles were excluded.

Evaluation of Bias

The bias of RCTs included in the systematic review was assessed using the Cochrane’s tool for assessing risk of bias [23], whereas the risk of bias in interventional cohort studies was assessed using ROBINS-I [25]. The evaluation was made by 2 independent assessors to ensure validity (LC1 and LC2). Disagreements were resolved by discussion with a consensus decision or by the decision of another author (XR). In case of duplicates, or multiple reports of a primary study, only the report with the most complete data set was included.

Data Extraction

Data extraction was performed independently by two reviewers (LC1 and LC2). Any discrepancies were resolved by discussion or by a third investigator (XR). All studies included in the meta-analysis had to be either RCTs or cohort studies. The prespecified primary outcomes were wound healing; the secondary outcomes were all-cause mortality, amputation, number of consultations, and patient experience.

Statistical Analysis

A meta-analysis was performed using Stata 12.0 (Stata Corp) and RevMan 5.3 (The Nordic Cochrane Centre, The Cochrane Collaboration). Hazard ratio (HR) and associated statistics were either extracted directly from articles or estimated from Kaplan–Meier curves [26]. Where sufficient data were available, a meta-analysis was conducted with a primary outcome of wound healing. The HRs for the healed wounds, with a value >1 favoring TM, were combined using a random-effect model (DerSimonian and Laird method). Pooled risk ratios (RRs), 95% CIs, and P values were estimated for endpoints in wound healing, all-cause mortality, amputation, and number of consultations, and patient experience.

Methods

Eligible Criteria

The systematic review and meta-analysis were conducted in accordance with recommendations by the Cochrane Collaboration [23] and the PRISMA statement [24]. Inclusion criteria were randomized controlled trials (RCT) and interventional cohort studies that investigated the use of TM follow-up in the community care in collaboration with specialists in wound centers with a comparator of no TM. Only articles that investigated chronic wounds, such as diabetic foot ulcers, stasis ulcer, pressure ulcers, and nonhealing surgical wound, were included. Exclusion criteria included case reports, editorials, letters, animal studies, case–control studies, and self-control studies. Non-English articles were excluded.

Search Strategy and Study Selection

A systematic search of databases (PubMed/MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials [CENTRAL]) was conducted to identify eligible studies published from inception to June 12, 2019. The reference lists of all identified articles and reviews were searched for potentially eligible studies. Only published articles were included. The search strategy is available in Multimedia Appendix 1. Two investigators selected studies independently (LC1 and WG). Disagreements were resolved by discussion with a consensus decision or by the decision of another author (XR). In case of duplicates, or multiple reports of a primary study, only the report with the most complete data set was included.
care model in the control arm, the studies were divided into 2 subgroups. In the community-based model, patients in the control arm received mainly routine wound care by general nurses in community or rural areas, who might not have enough expertise in wound care. In this model, patients might not receive standard care of wound management. By contrast, in the wound center-based model, patients received regular standard outpatient follow-up in the wound center. Sensitivity analysis was conducted by investigating the difference between random and fixed effects model as to effective measures. $P$ values $\leq 0.05$ were considered statistically significant.

In addition, we tested the hypothesis of noninferiority of TM follow-up for the primary efficacy outcomes. We adopted a $\Delta=0.15$ as margin of minimum clinically important differences [27,28]. For the primary efficacy outcomes, noninferiority of TM was demonstrated when the lower boundary of the 95% CI was greater than 0.85, which meant that TM could retain at least 85% of the effect of the conventional standard care.

### Results

A total of 12,007 potential studies were identified by the systematic search. Of these, 58 studies were selected for full review. Ultimately, 12 trials met the inclusion criteria, comprising a total of 3913 patients. Of these 5 were cluster RCTs [19-21,29,30], 1 was an RCT [22], and 6 were cohort studies [17,18,27,31-33]. The results of the study selection are shown in Multimedia Appendix 2.

### Study Characteristics

The selected studies [17-22,27,29-33] were published within 15 years. All studies included outpatients or patients in nursing homes. A total of 8 trials [17-19,27,29-32] treated patients with chronic wounds due to different etiologies (mixed wounds), 3 trials [20,22,33] treated patients with diabetic foot ulcers, and 1 study [21] enrolled patients with pressure ulcers. Whereas most studies included any severity of ulcers, one study [20] excluded patients with prior ulcers, which lead to a low proportion of severe ulcers. The method of TM delivery varied between studies. Four studies used tablet or mobile phones programmed with apps, such as Skype and specialized interactive systems [17,20,29,31], whereas other studies used email, telephone, and videoconferencing to facilitate the implementation of TM using specialized systems. However, all methods used included images of wound. The TM consultation specialists included specialized wound care nurses [20,21,32], wound care physicians [17,19,22,27,30,33], podiatrists [20], dermatologists [18,31], and a multidisciplinary team [21]. Wound care varied, particularly in the control group, because of the clinical heterogeneity of TM organization between studies, with 8 of these being community-based models [17-19,21,29-32], and the other 4 wound center-based models [20,22,27,33]. The treatment method and efficacy might have some differences between the two kinds of treatment models. Details of all 12 studies are summarized in Table 1.

### Risk of Bias

The result of assessment of risk of bias is presented in Multimedia Appendices 3-5. For RCTs, because allocation concealment and blinding would not seriously influence the selection of patients and the measurement of outcomes, there was no obvious bias in these two fields. The source of bias for RCTs mainly resulted from uneven baseline characteristics [19,20,29]. For cohort studies, the assessment of risk of bias using ROBINS-I showed that 2 studies [18,32] have moderate risk of bias which demonstrated a sound evidence for a nonrandomized study, whereas 4 studies [17,27,31,33] had serious risk of bias which demonstrated that the studies had some important problems.
### Table 1. Characteristics of the clinical trials included

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Wound etiology</th>
<th>No. of patients</th>
<th>Treatment strategy&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Control arm treatment location</th>
<th>Follow-up (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith-Strøm et al [20]</td>
<td>Norway</td>
<td>Diabetic foot ulcer</td>
<td>94</td>
<td>Via a web-based ulcer record and phone at least weekly; and outpatient consultation every 6 weeks</td>
<td>University hospital outpatient clinic</td>
<td>12</td>
</tr>
<tr>
<td>Stern et al [21]</td>
<td>Canada</td>
<td>Pressure ulcer</td>
<td>93</td>
<td>MDT consultation by email, telephone, or video</td>
<td>Usual care</td>
<td>12</td>
</tr>
<tr>
<td>Vowden and Vowden [29]</td>
<td>UK</td>
<td>Any etiology</td>
<td>17</td>
<td>Consultation by wound-assessment form and images weekly</td>
<td>Home nursing</td>
<td>6</td>
</tr>
<tr>
<td>Terry et al [19]</td>
<td>USA</td>
<td>Various etiology&lt;sup&gt;d&lt;/sup&gt;</td>
<td>62</td>
<td>Consultation by images weekly</td>
<td>Usual care</td>
<td>16</td>
</tr>
<tr>
<td>Santamaria et al [30]</td>
<td>Australia</td>
<td>Any etiology</td>
<td>50</td>
<td>Consultation by images and measurements every 2 weeks</td>
<td>Care from local wound care clinian</td>
<td>12</td>
</tr>
<tr>
<td>Rasmussen et al [22]</td>
<td>Denmark</td>
<td>Diabetic foot ulcer</td>
<td>193</td>
<td>Two consultations by telephone or online written consultations and one outpatient consultation cycle</td>
<td>Three outpatient consultation cycle</td>
<td>12</td>
</tr>
<tr>
<td>Le Goff-Pronost et al [31]</td>
<td>France</td>
<td>Any etiology</td>
<td>77</td>
<td>Via videoconference and photos once a week</td>
<td>Primary care</td>
<td>9</td>
</tr>
<tr>
<td>Gamus et al [27]</td>
<td>Israel</td>
<td>Any etiology</td>
<td>277</td>
<td>Via videoconference</td>
<td>Outpatient clinic</td>
<td>35</td>
</tr>
<tr>
<td>Wickström et al [17]</td>
<td>Sweden</td>
<td>Any etiology</td>
<td>100</td>
<td>Video consultation</td>
<td>Primary care</td>
<td>24</td>
</tr>
<tr>
<td>Bergersen et al [32]</td>
<td>Norway</td>
<td>Any etiology</td>
<td>32</td>
<td>Via wound support network every 4 weeks</td>
<td>Primary home care</td>
<td>3</td>
</tr>
<tr>
<td>Zarchi et al [18]</td>
<td>Denmark</td>
<td>Any etiology&lt;sup&gt;e&lt;/sup&gt;</td>
<td>50</td>
<td>Via a web-based program at a minimum of every second week</td>
<td>Home-care nursing</td>
<td>12</td>
</tr>
<tr>
<td>Wilbright et al [33]</td>
<td>USA</td>
<td>Diabetic foot ulcer</td>
<td>20</td>
<td>Via real-time interactive video weekly</td>
<td>Face-to-face consultation</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup>The TM arm received primary care in collaboration with specialists in wound center; patients in the control arm received follow-up by community nurses; in addition, patients in the wound center-based model received treatment at wound center.

<sup>b</sup>TM: telemedicine.

<sup>c</sup>MDT: multidisciplinary teams (comprising 2 enterostomal nurses and 1 certified wound-care nurse, or hospital-based wound-expert team).

<sup>d</sup>Nonhealing surgical wound, stasis ulcer, pressure ulcer.

<sup>e</sup>Surgical wounds, pressure ulcers, and cancer wounds excluded.

**Wound Healing**

Five studies [17,18,20-22] reported data on time to healing. Four studies [18,20-22] directly reported the effect measures HR and CI; however, HR was estimated from the Kaplan–Meier curve in the other remaining study [17]. Overall, TM appeared to demonstrate significant improvement in wound healing (HR 1.40, 95% CI 1.10–1.79; \(P=0.01; \, \hat{I}^2=60.6\%\); Figure 1). \(\hat{I}^2\) statistic of HR was 60.6% (\(P=0.04\), consistent with moderate heterogeneity of the analysis. In analysis stratified by study design, there was a trend in favor of TM in RCTs (HR 1.16, 95% CI 0.96–1.39; \(P=0.13; \, \hat{I}^2=0.0\%\)). The lower boundary of HR in RCTs was higher than 0.85, and the criteria of noninferiority of TM were met for wound healing. In addition, statistical difference was demonstrated in cohort studies demonstrating that TM had a decreased risk of allowing unhealed ulcers (HR 1.74, 95% CI 1.43–2.12; \(P<0.001; \, \hat{I}^2=0.0\%\)). We conducted a subgroup analysis of RCTs to investigate whether a different model of TM organization would result in different clinical effects. The studies that adopted the wound center–based model had a pooled HR of 1.13 (95% CI 0.93–1.37; \(P=0.22; \, \hat{I}^2=0.0\%\)), and the criteria of noninferiority of TM were
met (Multimedia Appendix 6). The 2 cohort studies were both community-based models.

Eight studies [17-22,29,31] reported a wound healing rate of around 1 year, including 5 RCTs and 3 cohort studies. One RCT [19] was not included in the quantitative synthesis because of uneven distribution of severity of wounds among groups. The pooled data of RCTs comparing TM and control showed no statistically significant difference in wound healing (RR 1.21, 95% CI 0.96-1.53; \( P=11 \); \( I^2=88.0\% \); Figure 2). This finding is consistent in RCTs (RR 1.05, 95% CI 0.89-1.23; \( P=15 \); \( I^2=45.2\% \)) and cohort studies (RR 1.32, 95% CI 0.91-1.91; \( P=56 \); \( I^2=85.2\% \)). However, these statistical trends were in favor of TM, and the criteria of noninferiority of TM were met. Two studies [32,33] reported wound healing at 3 months. Although there was a trend in favor of the TM group, no statistically significant difference between TM and control on wound healing at 3 months was revealed (RR 1.24, 95% CI 0.47-3.3; \( P=67 \); \( I^2=80.8\% \); Figure 3).

**Figure 1.** The effect of telemedicine on wound healing. HR: hazard ratio.

<table>
<thead>
<tr>
<th>Study</th>
<th>HR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Randomized trials</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stern et al [21]</td>
<td>1.48 (0.79-2.78)</td>
<td>10.66</td>
</tr>
<tr>
<td>Smith-Strøm et al [20]</td>
<td>1.16 (0.85-1.59)</td>
<td>22.92</td>
</tr>
<tr>
<td>Rasmussen et al [22]</td>
<td>1.11 (0.87-1.42)</td>
<td>26.87</td>
</tr>
<tr>
<td>Subtotal (( I^2 = 0.0% , P = .13 ))</td>
<td>1.16 (0.96-1.39)</td>
<td>60.45</td>
</tr>
<tr>
<td><strong>Cohort studies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wickström et al [17]</td>
<td>1.70 (1.38-2.08)</td>
<td>29.24</td>
</tr>
<tr>
<td>Zarchi et al [18]</td>
<td>2.19 (1.15-4.17)</td>
<td>10.31</td>
</tr>
<tr>
<td>Subtotal (( I^2 = 0.0% , P &lt; .001 ))</td>
<td>1.74 (1.43-2.12)</td>
<td>39.55</td>
</tr>
<tr>
<td>Overall (( I^2 = 60.6% , P = .01 ))</td>
<td>1.40 (1.10-1.79)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

*Note: Weights are from random effects analysis*
Figure 2. The effect of telemedicine on wound healing around 1 year. RR: risk ratio.

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stern et al [21]</td>
<td>1.24 (0.84-1.83)</td>
<td>12.98</td>
</tr>
<tr>
<td>Vowden and Vowden [29]</td>
<td>5.82 (0.89-38.20)</td>
<td>1.45</td>
</tr>
<tr>
<td>Smith-Ström et al [20]</td>
<td>1.05 (0.90-1.22)</td>
<td>18.52</td>
</tr>
<tr>
<td>Rasmussen et al [22]</td>
<td>0.97 (0.86-1.10)</td>
<td>19.06</td>
</tr>
<tr>
<td>Subtotal</td>
<td>1.05 (0.89-1.23)</td>
<td>52.02</td>
</tr>
<tr>
<td>Cohort studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le Goff-Pronost et al [31]</td>
<td>0.92 (0.69-1.22)</td>
<td>15.55</td>
</tr>
<tr>
<td>Wickström et al [17]</td>
<td>1.58 (1.43-1.75)</td>
<td>19.41</td>
</tr>
<tr>
<td>Zarchi et al [18]</td>
<td>1.56 (1.06-2.29)</td>
<td>13.02</td>
</tr>
<tr>
<td>Subtotal</td>
<td>1.32 (0.91-1.91)</td>
<td>47.98</td>
</tr>
<tr>
<td>Overall</td>
<td>1.21 (0.96-1.53)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

Figure 3. The effect of telemedicine on wound healing around 3 months. RR: risk ratio.

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergersen et al [32]</td>
<td>1.97 (0.84-4.60)</td>
<td>41.32</td>
</tr>
<tr>
<td>Wilbright et al [33]</td>
<td>0.89 (0.74-1.08)</td>
<td>58.68</td>
</tr>
<tr>
<td>Overall</td>
<td>1.24 (0.47-3.30)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

One study [30] revealed a positive healing rate of 6.8% per week, whereas controls had a negative rate of –4.9% per week ($P=.01$). In this study, no exact number of healed wounds and time to healing were provided, and therefore, the study was not considered in the quantitative analysis.

Another study [27] adopted positive outcomes (indicating at least 50% ulcer closure) as the primary outcome, so it was not included in the quantitative analysis. In this study, equality of TM and face-to-face methods was assessed using 2 one-sided noninferiority tests (WinPepi and the Westlake–Schuirmann method), and the noninferiority of TM was demonstrated within the $\Delta=0.15$ range limits and 80% statistical power.

In the trial with uneven distribution of severity and type of wounds among groups [19], the TM group had significantly larger wound size ($P=.03$) and more severe pressure ulcers. Although wounds in the TM group took longer time to heal and required more resources, it seems as though a greater change in size for pressure ulcers and other wounds occurred.
All-Cause Mortality

Eight studies [17,18,20-22,29-31] reported all-cause mortality around 1 year. Pooled data revealed no significant difference in mortality rate between the TM and control groups (RR 1.03, 95% CI 0.47-2.24; \( P=0.94; I^2=62.6\%\); Figure 4). This finding is consistent in RCTs (RR 0.92, 95% CI 0.30-2.83; \( P=0.89; I^2=59.5\%\)) and cohort studies (RR 1.29, 95% CI 0.44-3.75; \( P=0.64; I^2=51.4\%\)). Subgroup analysis of RCTs reveals there was a statistically significant decreased risk of all-cause mortality in patients receiving TM in the community-based model (RR 0.39, 95% CI 0.19-0.79; \( P=0.01; I^2=0.0\%\)); however, no statistical difference in mortality in the wound center–based model was demonstrated (RR 2.25, 95% CI 0.28-18.13; \( P=0.45; I^2=67.9\%\); Multimedia Appendix 7).

Figure 4. The effect of telemedicine on all-cause mortality. RR: risk ratio.

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized trials</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stern et al [21]</td>
<td>0.33 (0.15-0.72)</td>
<td>19.51</td>
</tr>
<tr>
<td>Vowden and Vowden [29]</td>
<td>1.59 (0.19-13.15)</td>
<td>8.67</td>
</tr>
<tr>
<td>Smith-Strom et al [20]</td>
<td>0.94 (0.28-3.12)</td>
<td>15.28</td>
</tr>
<tr>
<td>Rasmussen et al [22]</td>
<td>7.50 (0.95-59.39)</td>
<td>8.91</td>
</tr>
<tr>
<td>Santamaria et al [30]</td>
<td>0.29 (0.01-6.88)</td>
<td>4.80</td>
</tr>
<tr>
<td>Subtotal ((I^2 = 59.5%, P = .89))</td>
<td>0.92 (0.30-2.83)</td>
<td>57.18</td>
</tr>
<tr>
<td>Cohort studies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le Goff-Pronost et al [31]</td>
<td>2.79 (0.65-11.95)</td>
<td>13.07</td>
</tr>
<tr>
<td>Wickström et al [17]</td>
<td>1.57 (0.88-2.81)</td>
<td>21.28</td>
</tr>
<tr>
<td>Zarchi et al [18]</td>
<td>0.20 (0.02-1.72)</td>
<td>8.47</td>
</tr>
<tr>
<td>Subtotal ((I^2 = 51.4%, P = .64))</td>
<td>1.29 (0.44-3.75)</td>
<td>42.82</td>
</tr>
<tr>
<td>Overall ((I^2 = 62.6%, P = .94))</td>
<td>1.03 (0.47-2.24)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

Amputation

Three RCTs [20,22,30] reported amputation around 1 year. There was a statistically decreased risk of amputation in patients receiving TM follow-up care compared with conventional care (RR 0.45, 95% CI 0.29-0.71; \( P=0.001; I^2=0.0\%\); Figure 5).
Figure 5. The effect of telemedicine on amputation. RR: risk ratio.

<table>
<thead>
<tr>
<th>Study</th>
<th>RR (95% CI)</th>
<th>Weight (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith-Stoer et al [20]</td>
<td>0.43 (0.17-1.09)</td>
<td>24.57</td>
</tr>
<tr>
<td>Rasmussen et al [22]</td>
<td>0.50 (0.29-0.86)</td>
<td>70.58</td>
</tr>
<tr>
<td>Santamaria et al [30]</td>
<td>0.14 (0.02-1.14)</td>
<td>4.85</td>
</tr>
<tr>
<td>Overall ( (I^2 = 0.0%, \ P = .001) )</td>
<td>0.45 (0.29-0.71)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

NOTE: Weights are from random effects analysis

Number of Consultations

Three studies reported on the number of consultations. One RCT [20] revealed that mean consultations at the outpatient clinic per month in the TM group were not statistically different from conventional standard care of wounds (2.0 [SD 1.9] vs 2.5 [SD 3.0]). One cohort study also demonstrated no significant difference between TM and face-to-face follow-up (7.74 [SD 6.79] vs 9.18 [11.05]; \( P = .20 \) [27]. These two studies adopted a wound center-based model. Another cohort study [32], which adopted a community-based model, showed that the TM group had fewer (mean) appointments at the hospital compared with the control group (1.37 vs 2.33, \( P < .001 \)).

Patient Satisfaction

One study [20] evaluated patient satisfaction using the Generic Short Patient Experiences Questionnaire, and revealed that there was no statistically significant difference between the two groups (mean difference 0.07, 95% CI –0.10 to 0.24).

Economic Evaluation

A total of 4 studies [19, 21, 22, 32] mentioned a cost analysis of TM intervention versus control. A detailed economic analysis of 1 study [22] was published in another article [34], which found that the TM follow-up was US $2300 less per patient compared with standard care; however, the difference was not statistically significant (\( P = .42 \)). Another 2 studies [21, 32] also revealed reduced cost. One study [19] revealed a higher mean total cost per patient because of larger and more severe ulcers in the TM group.

Sensitivity Analysis

The sensitivity analysis is presented in Multimedia Appendix 8. All the efficacy outcomes were consistent between random and fixed models, except for wound healing around 1 year of cohort studies. We adopted the most conservative efficacy outcome (RR) using the random model, which still met the criteria of noninferiority of TM.

Discussion

Principal Results

In this review, we included 6 RCTs and 6 cohort studies comprising 3913 patients to evaluate the effect of TM in chronic wound management. We adopted both HR and RR to evaluate the effect of TM. In RCTs, we observed no significant differences in the primary clinical outcome efficacies HR and RR around 1 year, and noninferiority criteria were met. In cohort studies, the outcome efficacy HR was in favor of TM, whereas the efficacy RR around 1 year was not significantly different between TM and conventional standard of care of chronic wounds. Overall, these results showed that TM was noninferior to conventional standard of care. In terms of mortality, TM was not significantly different from control in both RCTs and cohort studies. A decreased risk of amputation was observed in patients receiving TM. A few studies performed qualitative analysis on the number of consultations, patient satisfaction, and economic evaluation with the results showing that TM was not worse than conventional standard care of chronic wounds. Therefore, TM seems to be a safe and effective method in the management of chronic wounds.

We carefully observed the difference of primary outcome between RCTs and cohort studies and found that the enrolled studies in RCTs were mainly wound center-based models, whereas those in cohort studies were all community-based models. Therefore, we speculate that the organization model may have a great influence on the effect of TM on wound healing; in particular, the community-based model may benefit from the implementation of TM. A possible explanation may be that in the community-based model, TM allowed patients in remote and rural settings easier access to multidisciplinary management which has been demonstrated to be an effective and efficient way of chronic wound management [35, 36].

Subgroup analysis of RCTs suggested that in a community-based model, patients in the TM group have a decreased risk of mortality. Both the positive primary outcome in cohort studies...
and the decreased mortality are in favor of TM in the community-based model. This demonstrates that it is promising to take advantage of TM in community or remote rural areas.

Comparison With Prior Studies

To our knowledge, this review included the largest number of patients with different types of wounds. The results of this review coincide with the systematic review by Tchero et al [16], which included 2 studies to investigate the effectiveness of TM in diabetic foot ulcer management. The authors found that patients in the TM and control groups had similar healing time (43 vs 45 days; P=.83) as well as similar ulcer healing rate (odds ratio 0.86, 95% CI 0.57-1.33; P=.53). A 10-year study of 5795 patients in France also provided an example of how TM might be of benefit in wound care [37].

Although similar mortality rates were revealed between TM and conventional care of chronic wound, the result of a well-designed RCT [22] revealed a higher mortality in the TM group (HR 8.68, 95% CI 6.93-10.88; P=.0001). In the authors’ opinion, the dependence of TM on secondhand information from a nurse could have caused some vital information to be missed. Therefore, it is worth noting that the severity of wounds and other commodities should be taken into consideration when health care participants are considering the use of TM to manage chronic wounds.

For the first time, we learned that the difference between the community- and wound center–based wound management models might seriously influence the effect of TM. A subgroup analysis was conducted to clarify the difference. Differing results between these two models indicate that it is prudent to understand the management model before interpreting the studies on TM in chronic wound management.

Limitations

This study has several limitations. First, we only searched 3 databases and did not include non-English literature. Although we tried to identify articles from reference lists of other reviews, it is possible that some studies in other databases or published in other language were overlooked. Second, we included RCTs and interventional cohort studies in the analysis, and thus, a potential source of bias might be introduced. Third, several studies [29,32,33] were deemed to be too small to be of statistical significance. Furthermore, in 1 trial [21], a large proportion of censored participants (107/201, 53%) could reduce statistical significance. Furthermore, in 1 trial [21], a large proportion of censored participants (107/201, 53%) could reduce statistical significance. Fourth, there were obvious variations between studies regarding number of participants, wound etiologies and degree of severity, and implementation of TM, such as video consultation, telephone, email, and picture transmission. Finally, with the ascent of information age, there has been a growing interest in TM. Researchers might tend to publish positive results. All these variations might contribute to bias.

Implications for Future Studies and Clinical Practices

First, for RCTs, although blinding of outcome measurement would not seriously influence the results, nonblinding of participants might bias the effect. In future studies, more importance should be assigned to the blinding of participants and health care providers. For example, all participants can receive treatment/suggestions via TM, but the information would not be sent to wound center specialists. In this way, performance bias could be reduced to a minimum.

Second, subgroup analysis indicates that TM in the community-based model is superior to standard primary care of chronic wounds by presenting with better outcomes and less mortality. Therefore, it is promising to take advantage of TM in community or remote rural areas. However, the number of studies in this aspect is limited and most studies are cohort studies. Thus, in future studies, well-designed, large-scale RCTs should be performed to verify the effect of TM in the community-based model.

Third, subgroup analysis indicates that TM in the wound center-based model is similar to standard of care of chronic wounds. Therefore, it is necessary to investigate whether TM can have better performance in other aspects. Only a few studies showed that TM was not worse than conventional standard of care regarding number of consultations, patient satisfaction, and economic evaluation. Thus, in future studies, these aspects can be included in the design of trials to investigate the effect of TM.

Finally, results in this systematic review and meta-analysis also shed some light on clinical practices. If health care practitioners would like to use TM to improve wound healing, they do not have to worry about delayed wound healing. For those patients who lived far away from wound center, TM can provide appropriate wound management.

Conclusions

TM is noninferior to conventional standard care of chronic wounds. TM might be a prosperous method for improving outcomes of patients living in remote or rural areas. However, owing to the relatively low quality of evidence, well-designed and adequately powered RCTs are further needed to confirm the role of TM.

Acknowledgments

This study was supported by the Health Medical Big Data Application and Innovation Project in Sichuan (Grant No. 2018gfw001), National Science and Technology Major Project (Grant No. 2017ZX09304023), National Key R&D Program of China (Grant No. 2017YFC1309605), Science and Technology Bureau of Sichuan Province (Grant Nos 2018SZ0239 and 2018JY0608), and Science and Technology Bureau of Chengdu city (Grant No. 2017-CY02-00028-GX). We thank Dr Aristidis Veves and Mr Navin Jayaswal for their help in language modification.
Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 27 KB - mhealth_v8i6e15574_app1.pdf]

Multimedia Appendix 2
Flow chart for study selection.
[PDF File (Adobe PDF File), 64 KB - mhealth_v8i6e15574_app2.pdf]

Multimedia Appendix 3
Risk of bias graph of RCTs.
[PDF File (Adobe PDF File), 238 KB - mhealth_v8i6e15574_app3.pdf]

Multimedia Appendix 4
Risk of bias summary of RCTs.
[PDF File (Adobe PDF File), 282 KB - mhealth_v8i6e15574_app4.pdf]

Multimedia Appendix 5
Risk of bias of cohort studies by the use of ROBINS-I.
[PDF File (Adobe PDF File), 478 KB - mhealth_v8i6e15574_app5.pdf]

Multimedia Appendix 6
Subgroup analysis of telemedicine on wound healing.
[PNG File, 56 KB - mhealth_v8i6e15574_app6.PNG]

Multimedia Appendix 7
Subgroup analysis of telemedicine on all-cause mortality.
[PNG File, 82 KB - mhealth_v8i6e15574_app7.PNG]

Multimedia Appendix 8
Results of sensitivity analysis.
[PDF File (Adobe PDF File), 269 KB - mhealth_v8i6e15574_app8.pdf]

References
8. Henneberg E. Telemedical wound assessment on the way to large scale deployment in Denmark. EWMA 2014;14:43-46.


Abbreviations

HR: hazard ratio
RCT: randomized controlled trials
RR: risk ratio
TM: telemedicine

©Lihong Chen, Lihui Cheng, Wei Gao, Dawei Chen, Chun Wang, Xingwu Ran. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 25.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Enhanced Patient Education for Colonic Polyp and Adenoma Detection: Meta-Analysis of Randomized Controlled Trials

Xu Tian*, MSN; Ling-Li Xu*, BSc; Xiao-Ling Liu*, BSc; Wei-Qing Chen, MD
Chongqing University Cancer Hospital, Chongqing, China
*these authors contributed equally

Corresponding Author:
Wei-Qing Chen, MD
Chongqing University Cancer Hospital
181 Hanyu Road
Shapingba District
Chongqing, 400030
China
Phone: 86 1822275023
Email: CQCH_ChenWQ@163.com

Abstract

Background: To improve patients’ comprehension of bowel preparation instructions before colonoscopy, enhanced patient education (EPE) such as cartoon pictures or other visual aids, phone calls, mobile apps, multimedia education and social media apps have been proposed. However, it is uncertain whether EPE can increase the detection rate of colonic polyps and adenomas.

Objective: This meta-analysis aimed to evaluate the efficacy of EPE in detecting colonic polyps and adenomas.

Methods: We searched PubMed, EMBASE, and Cochrane Central Register of Controlled Trials from their inception to June 2019 for the identification of trials comparing the EPE with standard patient education for outpatients undergoing colonoscopy. We used a random effects model to calculate summary estimates of the polyp detection rate (defined as the number of patients with at least one polyp divided by the total number of patients undergoing selective colonoscopy), adenoma detection rate (defined as the number of patients with at least one adenoma divided by the total number of patients undergoing selective colonoscopy), advanced adenoma detection rate (defined as the number of patients with at least one advanced adenoma divided by the total number of patients undergoing selective colonoscopy), sessile serrated adenoma detection rate (defined as the number of patients with at least one sessile serrated adenoma divided by the total number of patients undergoing selective colonoscopy), cancer detection rate (defined as the number of patients with at least one cancer divided by the total number of patients undergoing selective colonoscopy), or adenoma detection rate - plus (defined as the number of additional adenomas found after the first adenoma per colonoscopy). Moreover, we conducted trial sequential analysis (TSA) to determine the robustness of summary estimates of all primary outcomes.

Results: We included 10 randomized controlled trials enrolling 4560 participants for analysis. The meta-analysis suggested that EPE was associated with an increased polyp detection rate (9 trials; 3781 participants; risk ratio [RR] 1.19, 95% CI 1.05-1.35; P<.05; I²=42%) and adenoma detection rate (5 trials; 2133 participants; RR 1.37, 95% CI 1.15-1.64; P<.001; I²=0%), which were established by TSA. Pooled result from the inverse-variance model illustrated an increase in the sessile serrated adenoma detection rate (3 trials; 1248 participants; odds ratio 1.76, 95% CI 1.22-2.53; P<.05; I²=0%). One trial suggested an increase in the adenoma detection rate - plus (RR 4.39, 95% CI 2.91-6.61; P<.001). Pooled estimates from 3 (1649 participants) and 2 trials (1375 participants) generated no evidence of statistical difference for the advanced adenoma detection rate and cancer detection rate, respectively.

Conclusions: The current evidence indicates that EPE should be recommended to instruct bowel preparation in patients undergoing colonoscopy because it can increase the polyp detection rate, adenoma detection rate, and sessile serrated adenoma detection rate. However, further trials are warranted to determine the efficacy of EPE for advanced adenoma detection rate, adenoma detection rate - plus, and cancer detection rate because of limited data.

(JMIR Mhealth Uhealth 2020;8(6):e17372) doi:10.2196/17372
KEYWORDS
colonoscopy; bowel preparation; patient education; polyp detection rate; adenoma detection rate; meta-analysis

Introduction

Background
Colorectal cancer (CRC) is the third most common cancer and the second cause of cancer-related mortality among both sexes worldwide, with 1.8 million new cases and 0.88 million deaths in 2018 [1]. Colonoscopy is recommended as the principal approach for decreasing CRC incidence and associated mortality by detecting and then removing the precancerous lesions [2-5]. However, adequate bowel preparation is the prerequisite for a successful colonoscopy [6]. Evidence revealed that inadequate bowel preparation was associated with increased risk of missing colonic lesions, prolonged procedural time, and lower cecal intubation rate [7,8]. Issued data suggested approximately 18% to 30.5% of inadequate bowel preparation in patients undergoing colonoscopy [9-11]. Therefore, it is particularly urgent to apply an effective intervention to improve the quality of bowel preparation [12].

Previous studies have determined various factors that were linked to the quality of bowel preparation, such as the type of diet restriction, type of colon cleansing solutions ingested, methods of ingesting the solution, and patient’s adherence to the solution [10,13-15]. Adequate comprehension of details of instructions is a major contributor to the quality of colon cleansing because bowel preparation is very complex [16]. Patients usually receive written booklet and/or verbal instructions from professionals before colonoscopy for bowel preparation and dietary restriction, which are defined as standard patient education [2]. However, the effect of standard patient education on bowel preparation is not enough [10]. To improve the patient’s comprehension of bowel preparation instructions, enhanced patient education (EPE; such as cartoon pictures, phone calls, mobile apps, and social media apps) has been proposed and then tested [2]. So far, several meta-analyses have evaluated the efficacy of EPE in improving the quality of bowel preparation and demonstrated an improvement [2,16-19]. However, evidence revealed that adequate bowel preparation provides good colonoscopy vision and thus increases the detection rate of colonic polyps and adenomas [20-23]. The fact that 2 meta-analyses evaluated the efficacy of EPE interventions to detect polyps and both did not find significant differences is discouraging [2,17]. As a result, the magnitude of benefit of EPE interventions in detecting colonic polyps and adenomas remains uncertain. It is noteworthy that recently, several randomized controlled trials (RCTs) reporting conflicting results have been published. More importantly, as most CRCs transform from polyps and adenomas, early detection and then removal of premalignant colonic polyps and adenomas is crucial [19]. Thus, as one of the most important colonoscopy quality metrics, the detection of colonic polyps and adenomas should be primarily measured and evaluated [24].

Objective
The aim of this meta-analysis was to evaluate the efficacy of EPE interventions in detecting CRC precursor polyps and adenomas compared with standard patient education.

Methods

Methodological Standard
We conducted this meta-analysis according to the methods proposed by the Cochrane Collaboration [25] and reported the pooled estimates in accordance with the framework proposed by the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement [26]. There was no formal protocol for this meta-analysis.

Search Strategy
A systematic search was performed in PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials from their inception to June 2019 for the identification of relevant RCTs. All search strategies were built using Exploded Medical Subject Headings and the appropriate corresponding text words. Language and status of publication were not restricted. We have summarized the details of the full search strategy in Multimedia Appendix 1. We updated the search results on August 10, 2019. The bibliographies of previous meta-analyses and eligible studies were also manually checked to identify additional potentially eligible studies.

Study Selection
We used the following inclusion criteria to enroll any eligible studies in this study: (1) all adult participants aged more than 18 years who were instructed to receive elective outpatient colonoscopy, regardless of morning and afternoon colonoscopy; (2) the patients assigned in the study group were instructed with EPE regimes, and the ones enrolled in the control group were instructed with standard patient education regimes; (3) the eligible study design was RCTs; however, an abstract with sufficient information was also considered; and (4) studies published in English. Two investigators (XT and LX) independently searched citations; excluded duplicates; checked the titles and abstracts for eligibility; and then categorized the studies as included, excluded, or requiring further full-text assessment. We excluded duplicates with poor quality or relatively insufficient data. We also excluded conference abstracts without sufficient information. A third senior investigator (WC) was consulted for a final decision if there was any disagreement between the 2 investigators.

Data Extraction
Two independent investigators (XT and LX) were assigned to use a standardized Word (version 2013, Microsoft Office, Microsoft Corporation) table to extract essential data, and then, they completed the cross-checking of corresponding results. The following data were extracted: basic characteristics of eligible trials including leading author, publication year, country, and financial sources; risk of bias criteria based on the Cochrane
Collaboration risk of bias tool [27]; and clinical characteristics including age, sex, sample size, indication for colonoscopy, details of diet restriction and colon cleansing solutions, details of education interventions, and outcomes of interest. A third senior investigator (WC) was consulted for a final decision if there was any disagreement between the 2 investigators.

**Outcome Variables and Definitions**

We defined the colonic polyp detection rate (PDR) and adenoma detection rate (ADR) as the primary outcomes, which were defined as the number of patients with at least one polyp or adenoma divided by the total number of patients undergoing selective colonoscopy. We considered the advanced adenoma (defined as adenoma ≥10 mm) detection rate (AADR), sessile serrated adenoma detection rate (SSADR), and cancer detection rate (CDR) as secondary outcomes, which were defined as the number of patients with at least one advanced adenoma, sessile serrated adenoma, or cancer divided, respectively, by the total number of patients undergoing selective colonoscopy. We also considered ADR-plus, which was defined as the number of additional adenomas found after the first adenoma per colonoscopy [28], and the right and left colon polyp and ADR, which was defined as the number of patients with at least one right and left colon polyp and adenoma divided by the total number of patients undergoing selective colonoscopy, as secondary outcomes.

**Assessment of Risk of Bias**

We assigned 2 independent investigators (XT and XL) to appraise the risk of bias with the Cochrane risk of bias tool (the Cochrane Collaboration) [27]. An individual trial would be labeled as low, unclear, or high risk of bias according to the following criteria: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. A third senior investigator (WC) was consulted for a final decision if there was any disagreement between the 2 investigators. Following the recommendations proposed by the Cochrane Collaboration, a trial was considered a high-level trial when all key domains are rated as having a low risk of bias, a trial was considered a low-level trial when any one or more key domains are rated as having a high risk of bias, and otherwise, a trial was considered a moderate-level trial.

**Data Analysis**

We expressed summary estimates as a risk ratio (RR) or odds ratio (OR) with 95% CI. Heterogeneity was measured by the I² statistic, which describes the percentage of total variation across studies that is due to heterogeneity rather than chance [29]. We performed all statistical analyses using random effects model regardless of heterogeneity. In addition, subgroup analyses for the primary outcomes were conducted according to geographical regions (Western vs Asian) and indications (screening vs diagnostic vs mixed). We also conducted subgroup analysis for primary outcomes according to the administration method of ingesting colon cleans solutions (single dose vs split dose) because Zawaly et al [30] demonstrated that split dose compared with single dose was associated with increased detection of adenomas and advanced adenomas. P<.05 was considered statistically significant, except where otherwise specified. All statistical analyses were performed using Cochrane Review Manager (RevMan, version 5.3.5, 2014; the Nordic Cochrane Centre, the Cochrane Collaboration) [31]. For pooled estimates of SSADR, we used inverse-variance statistic due to various data reported in analyzed trials, and for remaining pooled estimates, we used the Mantel-Haenszel model. An RR or OR value greater than 1 indicates that there was a higher detection rate of the specified colonic polyp and adenoma. We planned to assess publication bias if any pooled group consisted of 10 or more trials [32].

**Trial Sequential Analysis**

The magnitude of efficacy of summary estimates from cumulative meta-analyses and the risk of type I error are susceptible to repetitive hypothesis test of accumulating scarce information [33]. Thus, trial sequential analysis (TSA), which has the potential of constantly adjusting the significance level and then drawing monitoring boundaries and calculating adjusted information size, was proposed to address issues faced by the traditional meta-analysis [34-36]. The conclusion is conclusive if the accumulative sample size is more than the adjusted information size and the Z-curve is across the trial sequential monitoring boundary or futility boundary. We conducted TSA to test the robustness of summary estimates of primary outcomes according to an alpha error of .05, a beta error of .20 (a power of 80%), and an anticipated intervention effect of 20% relative risk reduction using TSA version 0.9 beta (Copenhagen Trial Unit, Center for Clinical Intervention Research) [37,38].

**Results**

**Literature Search**

Figure 1 depicts the retrieval and selection of records. Initial search captured 588 records in 3 targeted databases. All records were imported to EndNote software (Thomson Reuters), and then, we deleted 110 duplicate records after running the finding duplicates function. We excluded additional 404 records after checking the titles and abstracts because of the following reasons: systematic review and meta-analysis and irrelevant to the analysis. We omitted 64 studies after carefully double-checking the full text in the remaining 74 studies because of the following reasons: systematic review and meta-analysis and irrelevant to the analysis. We. We omitted 64 studies after carefully double-checking the full text in the remaining 74 studies because of the following reasons: 9 articles investigated a topic unrelated to this study, 1 article was a letter to the editor, 3 articles were editorials, 1 article was a duplicate publication, 1 article was a comment on published article, 36 were conference abstracts without sufficient information, 9 articles did not report essential outcomes or data that were considered in our study, and 4 articles used ineligible study design. We, thus, included 10 eligible RCTs in the final meta-analysis after checking the full text for eligibility [9,39-47].
Study Characteristics

We document the details of basic characteristics of the 10 eligible RCTs in Table 1. All trials were reported between 2011 and 2018. The sample size in individual trial ranged from 94 to 969 (a total of 4560 participants). Of the 10 eligible RCTs, 3 [39,41,46] were from Western countries, including the United States [39,41] and Germany [46], and 7 [9,40,42-45,47] were from Asian countries, including China [9,43,47], Korea [40,45], and South Korea [42,44]. In total, 8 RCTs [9,39,41,43-46] were designed with two arms and remaining 2 RCTs [42,47] with three arms. The participants in 5 RCTs [39,40,42,44-45] received screening colonoscopy; in 4 RCTs [9,41,43,46] received screening, surveillance, or diagnostic colonoscopy; and in 1 RCT [47] received diagnostic colonoscopy. A total of 8 RCTs [9,40,44,46,47] described the details of diet restriction, and other 2 RCTs [39,45] did not report on the diet. Of the 10 included RCTs, 9 reported PDR as outcome [39-47], 5 reported ADR as outcome [9,41,42,46,47], 3 reported AADR as outcome [9,46,47], 3 reported SSADR as outcome [9,41,46], 2 reported CDR as outcome [9,43], and 1 reported ADR-plus as outcome [46].

Risk of Bias

Details of the risk of bias of individual trial are summarized in Table 2: 7 trials were rated as low level [39-42,44-46], 2 were moderate level [9,43], and 1 was high level [47]. A total of 8 trials appropriately generated randomization sequence [9,39,41-43,45-47], and 5 trials correctly conducted allocation concealment [9,42,43,46,47]. All trials [9,40-47] appropriately blinded endoscopist and reported anticipated outcomes.
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Sample size (SPE/EPE)</th>
<th>Sex (male/female; SPE/EPE)</th>
<th>Education strategies</th>
<th>Indications</th>
<th>Bowel cleansing regimen</th>
<th>Diet restriction</th>
<th>Start time of education</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calderwood et al (2011) [39]</td>
<td>United States</td>
<td>969 (492/477)</td>
<td>(205/287; 198/279)</td>
<td>Standard written pre-colonoscopy information</td>
<td>Screening colonoscopy</td>
<td>4 L PEG alone or plus bisacodyl</td>
<td>NR&lt;sup&gt;d&lt;/sup&gt;</td>
<td>NR</td>
<td>PDR&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cho et al (2017) [40]</td>
<td>Korea</td>
<td>142 (71/71)</td>
<td>(42/29; 42/29)</td>
<td>Verbal and written instructions</td>
<td>Smartphone app</td>
<td>Screening colonoscopy</td>
<td>Low residue</td>
<td>3 days before</td>
<td>PDR</td>
</tr>
<tr>
<td>Kang et al (2016) [9]</td>
<td>China</td>
<td>770 (383/387)</td>
<td>(191/192; 202/185)</td>
<td>Verbal and written instructions</td>
<td>WeChat</td>
<td>Mixed&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4 L PEG 4000 with split dose</td>
<td>Clear liquid</td>
<td>15 days before</td>
</tr>
<tr>
<td>Lee et al (2015) [42]</td>
<td>South Korea</td>
<td>394 (137/255)</td>
<td>(73/64; 155/98)</td>
<td>Verbal and written instructions</td>
<td>Telephone or SMS reminder</td>
<td>Screening colonoscopy</td>
<td>Low residue</td>
<td>2 days before</td>
<td>PDR and ADR</td>
</tr>
<tr>
<td>Liu et al (2013) [43]</td>
<td>China</td>
<td>605 (300/305)</td>
<td>(147/153; 160/145)</td>
<td>Verbal and written instructions</td>
<td>Telephone re-education</td>
<td>Mixed</td>
<td>2 L PEG plus ascorbic acid with split dose</td>
<td>Clear liquid</td>
<td>1 day before</td>
</tr>
<tr>
<td>Park et al (2016) [44]</td>
<td>South Korea</td>
<td>502 (252/250)</td>
<td>(167/85; 157/93)</td>
<td>Regular instruction</td>
<td>Educational video</td>
<td>Screening colonoscopy</td>
<td>Clear liquid</td>
<td>1 day before</td>
<td>PDR</td>
</tr>
<tr>
<td>Tae et al (2012) [45]</td>
<td>Korea</td>
<td>205 (103/102)</td>
<td>(71/32; 73/29)</td>
<td>Verbal and written instructions</td>
<td>Cartoon visual aids</td>
<td>Screening colonoscopy</td>
<td>PEG with split dose</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Wang et al (2018) [47]</td>
<td>China</td>
<td>384 (127/257)</td>
<td>(68/59; 149/108)</td>
<td>Verbal and written instructions</td>
<td>WeChat or SMS</td>
<td>Diagnostic colonoscopy</td>
<td>Clear liquid</td>
<td>2 days before</td>
<td>PDR</td>
</tr>
</tbody>
</table>

<sup>a</sup>SPE: standard patient education.<br/><sup>b</sup>EPE: enhanced patient education.<br/><sup>c</sup>PEG: polyethylene glycol.<br/><sup>d</sup>NR: not reported.<br/><sup>e</sup>PDR: polyp detection rate.<br/><sup>f</sup>ADR: adenoma detection rate.
Table 2. Details of quality assessment of eligible studies using the Cochrane risk of bias tool.

<table>
<thead>
<tr>
<th>Study</th>
<th>Random sequence generation</th>
<th>Allocation concealment</th>
<th>Blinding of participants and personnel</th>
<th>Blinding of outcome assessment</th>
<th>Incomplete outcome date</th>
<th>Selective reporting</th>
<th>Other bias</th>
<th>Overall level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calderwood et al (2011) [39]</td>
<td>Low risk</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Cho et al (2017) [40]</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Garg et al (2016) [41]</td>
<td>Low risk</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Lee et al (2015) [42]</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Tae et al (2012) [45]</td>
<td>Low risk</td>
<td>Unclear risk</td>
<td>Low risk</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low level</td>
</tr>
</tbody>
</table>

Primary Outcomes

Figure 2 depicts the summary results of primary outcomes. Meta-analysis based on a random effects model suggested an increase in the detection rate of polyps (9 trials; 35.55% [715/2011] of participants vs 30.23% [535/1770] of participants; RR 1.19, 95% CI 1.05-1.35; \( P=0.008; I^2=42\% \)) and adenomas (5 trials; 22.72% [271/1193] of participants vs 16.5% [155/940] of participants; RR 1.37, 95% CI 1.15-1.64; \( P<0.001; I^2=0\% \)) in patients undergoing EPE. Subgroup analyses indicated that EPE increased the PDR in Asian patients; in patients ingested solutions with split dose and single dose; and in mixed patients undergoing screening, surveillance, and diagnostic colonoscopy, and that EPE increased the ADR in all patients regardless of geographical regions and in patients ingested solutions with split dose. The summary of subgroup analyses is shown in Multimedia Appendix 2.

TSA suggested that the accumulative Z-curve crossed the trial sequential monitoring boundary for benefit after the eighth trial of PDR (Figure 3) and after the fourth trial of ADR (Figure 4), showing that currently, the cumulative evidence for PDR and ADR is conclusive.
Figure 2. Meta-analysis of the effect of EPE on PDR (A) and ADR (B). This pooled result indicated a statistical difference regarding PDR and ADR between EPE and SPE groups. The summary effect estimates (risk ratio, RR) for individual randomized controlled trial (RCT) are indicated by blue rectangles (the size of the rectangle is proportional to the study weight), with the black horizontal lines representing 95% CIs. The overall summary effect estimate (RR) and 95% CI are indicated by the black diamond below. EPE: enhanced patient education; SPE: standard patient education; PDR: polyp detection rate; ADR: adenomas detection rate; RR: risk ratio; and M-H: Mantel-Haenszel.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>EPE</th>
<th>Total</th>
<th>SPE</th>
<th>Total</th>
<th>Weight</th>
<th>Risk Ratio</th>
<th>M-H</th>
<th>Random</th>
<th>95% CI</th>
<th>Year</th>
<th>Risk Ratio</th>
<th>M-H</th>
<th>Random</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calderwood 2011</td>
<td>182</td>
<td>477</td>
<td>169</td>
<td>492</td>
<td>19.9%</td>
<td>0.99 [0.85, 1.17]</td>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tae 2012</td>
<td>55</td>
<td>102</td>
<td>53</td>
<td>98</td>
<td>13.5%</td>
<td>1.00 [0.77, 1.29]</td>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liu 2013</td>
<td>116</td>
<td>305</td>
<td>74</td>
<td>300</td>
<td>14.1%</td>
<td>1.54 [1.21, 1.97]</td>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee 2015</td>
<td>101</td>
<td>253</td>
<td>44</td>
<td>137</td>
<td>11.9%</td>
<td>1.24 [0.93, 1.66]</td>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garg 2016</td>
<td>23</td>
<td>48</td>
<td>16</td>
<td>46</td>
<td>5.5%</td>
<td>1.38 [0.84, 2.26]</td>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park 2016</td>
<td>34</td>
<td>250</td>
<td>34</td>
<td>252</td>
<td>6.5%</td>
<td>1.01 [0.65, 1.57]</td>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cho 2017</td>
<td>23</td>
<td>71</td>
<td>15</td>
<td>71</td>
<td>4.4%</td>
<td>1.53 [0.87, 2.69]</td>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang 2018</td>
<td>81</td>
<td>257</td>
<td>36</td>
<td>127</td>
<td>9.9%</td>
<td>1.11 [0.80, 1.55]</td>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walter 2018</td>
<td>100</td>
<td>248</td>
<td>74</td>
<td>247</td>
<td>14.2%</td>
<td>1.36 [1.06, 1.72]</td>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>211</td>
<td>577</td>
<td>177</td>
<td>700</td>
<td><strong>100.0%</strong></td>
<td>1.19 [1.05, 1.35]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total events: 715
Heterogeneity: $\tau^2 = 0.01; \chi^2 = 13.77, df = 8 (P = 0.09); I^2 = 42$
Test for overall effect: $Z = 2.67 (P = 0.008)$

Figure 3. Trial sequential analysis of PDR. A diversity-adjusted information size of 6356 patients was calculated using alpha=.05 (2-sided), beta=.20 (power 80%), an anticipated relative risk reduction of 20%, and an event proportion of 30.23% in the SPE arm. The TSA-adjusted 95% CI for a relative risk of 1.31 is 1.10 to 1.56 (random effects model [DL]). TSA illustrated that the required information size was not achieved (adjusted information size=6356), however, that the cumulative z curve crossed the trial sequential monitoring boundary for benefit, showing that currently cumulative evidence is conclusive. PDR: polyp detection rate; EPE: enhanced patient education; SPE: standard patient education; TSA: trial sequential analysis; DL: DerSimonian and Laird.
Figure 4. Trial sequential analysis of ADR. A diversity-adjusted information size of 3644 patients was calculated using alpha=.05 (2-sided), beta=.20 (power 80%), an anticipated relative risk reduction of 20%, and an event proportion of 16.49% in the SPE arm. The TSA-adjusted 95% CI for a relative risk of 1.37 is 1.15 to 1.64 (Random effects model [DL]). TSA illustrated that the required information size was not achieved (adjusted information size=3644), however, that the cumulative z curve crossed the trial sequential monitoring boundary for benefit, showing that currently cumulative evidence is conclusive. ADR: adenoma detection rate; EPE: enhanced patient education; SPE: standard patient education; TSA: trial sequential analysis; DL: DerSimonian and Laird.

Secondary Outcomes

Figure 5 delineated the summarized results of the secondary outcomes. Meta-analysis showed no significant difference in AADR (3 trials; 5.6% [50/892] of participants vs 3.8% [29/757] of participants; RR 1.54, 95% CI 0.67-3.55; \( P=0.31; I^2=57\%\) and CDR (2 trials; 1.59% vs 1.02%; RR 1.54, 95% CI 0.60-3.98; \( P=0.37; I^2=0\%\)), and a significant difference in SSADR based on inverse-variance model (3 trials; OR 1.76, 95% CI 1.22-2.53; \( P<0.05; I^2=0\%\)) between the EPE and standard patient education groups. Only 1 trial reported the ADR-plus, and adjusted estimate found a superior result in the EPE group (1 trial; RR 4.39, 95% CI 2.91-6.61; \( P<0.001\) [47]. One trial reported ADR according to segments of colon (right vs left) and did not find a significant difference between EPE and standard patient education interventions (right colon: RR 1.86, 95% CI 0.66-5.24 and left colon: RR 2.45, 95% CI 0.59-10.11) [41].

Publication Bias

Although the accumulated number of analyzed trials for all outcomes was less than 10, we also constructed the funnel plot for PDR because 9 trials were incorporated into this outcome. The funnel plot is not symmetrical, and the publication bias cannot be excluded (Figure 6).
Figure 5. Meta-analysis of the effect of EPE on AADR (A), SSADR (B), and CDR (C). The summary effect estimates (odds ratio) for individual randomized controlled trials are indicated by blue or red rectangles (the size of the rectangle is proportional to the study weight), with the black horizontal lines representing 95% CIs. The overall summary effect estimate (OR) and 95% CI are indicated by the black diamond below. EPE: enhanced patient education; SPE: standard patient education; PDR: polyp detection rate; ADR: adenomas detection rate; RR: risk ratio; M-H: Mantel-Haenszel; IV: inverse variance.

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>EPE Events</th>
<th>Total</th>
<th>Risk Ratio M-H Random 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kang 2016</td>
<td>9</td>
<td>20</td>
<td>1.27 [0.48, 3.38]</td>
</tr>
<tr>
<td>Walter 2018</td>
<td>20</td>
<td>24</td>
<td>1.00 [0.55, 1.80]</td>
</tr>
<tr>
<td>Wang 2018</td>
<td>21</td>
<td>27</td>
<td>5.19 [1.24, 21.78]</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>892</td>
<td>757</td>
<td>1.54 [0.67, 3.55]</td>
</tr>
<tr>
<td>Total events</td>
<td>50</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: Tau^2 = 0.30; Chi^2 = 4.60, df = 2 (P = 0.10); I^2 = 57%
Test for overall effect: Z = 1.01 (P = 0.31)

Figure 6. Funnel plot of PDR between the EPE and SPE groups. The vertical axis represents the standard error (SE) of effect size and x-axis indicates pooled risk ratio (RR). Symmetrical funnel plot indicates the absence of publication bias. PDR: polyp detection rate; EPE: enhanced patient education; SPE: standard patient education.
Discussion

Principal Findings

Adequate bowel preparation is a critical contributor to successful colonoscopy; however, the performance of traditional instructions of bowel preparation before colonoscopy in improving the quality of bowel preparation is not enough [10]. So, several enhanced patient instructions such as WeChat and SMS were developed to cover the shortcomings of the traditional instructions [2]. In this meta-analysis evaluating the detection rate of colonic polyps and adenomas, we found that EPE relatively increased PDR by 19% (95% CI 0.87-1.51) and ADR by 37% (95% CI 0.61/1,000), when compared with standard patient education before colonoscopy. Moreover, we also found that EPE was associated with increased SSADR (95% CI 1.22-2.53) and ADR-plus (RR 4.39, 95% CI 2.91-6.61). We found no evidence of statistical differences in AADR and CDR between the two groups. After performing a subgroup analysis, we found that EPE increased PDR by 21% in Asian patients and by 44% in mixed outliers, and ADR by 39% and 37% in Western and Asian patients undergoing colonoscopy, respectively. Although our study found an increased PDR, ADR, SSADR, and ADR-plus in the EPE group, the pure efficacy of EPE alone in the detection of colonic polyps and adenomas needs to be further investigated because all patients in the EPE group also received standard patient education.

Mechanism of Enhanced Patient Education

The EPE regimes were quite diverse in the 10 eligible trials, including visual aid, new visual aids, phone call, SMS, mobile apps, and multimedia education. We compared the essential characteristics of these EPE regimes with those of standard patient education (Multimedia Appendix 3). EPE has been shown to be effective in improving bowel preparation quality [2,19]. In this study, we further found that EPE increases the detection rate of colonic polyp and adenoma. EPE has several advantages including easy understanding of the education materials, easy access to the information of bowel preparation and dietary recommendations, more interactive approaches for seeking solutions to problems, or an additional approach to enhance correct memory compared with standard patient education [2]. After instructing patients with the EPE regime, the compliance of patients with instructions, including dietary recommendations and digestion of bowel preparation solutions, improves, which may be the possible reason for an increase in the colonic polyp and adenomas detection rate [2,9,42,43].

Comparison With Other Studies

To date, 5 systematic reviews and meta-analyses have been performed to comprehensively investigate the impact of EPE on the quality of bowel preparation compared with standard patient education [2,16-19]. However, only 2 of those 5 evaluated the PDR as a secondary outcome [2,17]. Chang et al [37] conducted a meta-analysis of 3 trials to evaluate the impact of EPE on PDR in patients undergoing colonoscopy and found no significant difference between the 2 groups (RR 1.14, 95% CI 0.87-1.51; \(I^2=79.1\%\)). However, there are some limitations that need to be considered in this meta-analysis. First, only 3 trials including 1779 patients were analyzed, which may cause summary estimates to be inflated and thereby limit the strength of the inference that can be drawn due to inadequate accumulated sample size (adjusted information size=6356) [48]. Therefore, the finding of this study may not be considered as definitive. Furthermore, this meta-analysis did not consider other important colonoscopy quality metrics such as ADR and adenomas per positive participant [49], which greatly reduced the relevance for clinical decision.

Guo et al [2] performed an updated systematic review and meta-analysis after including recent trials to evaluate the efficacy of EPE for bowel preparation before colonoscopy. In this study, PDR was also evaluated as a secondary outcome, and a pooled estimate based on 5 trials did not detect a significant difference in PDR between the 2 groups (OR 1.25, 95% CI 0.93-1.68; \(P=0.14\)). It must be noted, however, that there are also some flaws that need to be acknowledged in this study. First, although this study included more trials for final analysis [39,42,43,50,51], 2 studies that did not report appropriate data were inappropriately considered to be eligible [50,51]. As a result, the findings of this study must be interpreted cautiously. Similarly, this study did not also consider other important colonoscopy quality metrics [49].

Differences between our study and the 2 previous meta-analyses should be emphasized. In the 2 previous meta-analyses, limited number of eligible trials were accumulated for PDR, and since then, additional eligible trials with a high quality and large sample size have been published [9,40,41,44,46,47]. ADR is widely accepted as one of the objective colonoscopy quality metrics [52], and published evidence suggested that increased ADR was associated with decreased risk of post colonoscopy CRCs [53]. Moreover, ADR-plus was also suggested as one of the colonoscopy quality metrics with significant clinical relevance [28]. It is noted, however, that the previous meta-analyses only evaluated PDR, and other important quality metrics were not considered because of limited data, which reduced the clinical relevance of summary estimates. Our meta-analysis of 10 RCTs involving 4560 patients suggests that EPE is associated with an increase in PDR, ADR, ADR-plus, and SSADR. There were no significant differences in ADR and CDR. To test the robustness of summary results of primary outcomes, we performed a trial sequential meta-analysis to adjust the significance level and calculate the adjusted information size. Trial sequential analyses of both outcomes indicated that currently, cumulative evidences are conclusive. Moreover, subgroup meta-analyses found that EPE is only associated with increased PDR in Asian patients and ADR in all patients.

Strengths and Limitations of This Study

Strengths of this review include a comprehensive literature search and the inclusion of multiple types of precursor polyps as outcomes. Moreover, we also performed a trial sequential meta-analysis for PDR and ADR. However, limitations of this meta-analysis should be acknowledged. First, 2 trials with a three-arm design were included, and we simply combined the data reported in the two positive groups according to the methodology recommended by the Cochrane Collaboration.
However, it may be rational to use the network meta-analysis to compare all interventions if more eligible trials can be captured. Second, most trials analyzed in our meta-analysis assigned patients to follow different diet restrictions such as clear liquid and low-residue diets, but we did not perform subgroup analysis according to this condition because 2 trials did not introduce the details of diet restriction [39,45]. It is worth mentioning, however, that previous meta-analyses have demonstrated comparable efficacy between clear liquid and low-residue diets for bowel preparation before colonoscopy [54,55]. Third, our previous meta-analysis detected no significant difference between the low- and traditional-volume polyethylene glycol regimens in bowel preparation [56]. Thus, bowel preparation regimen was not considered to be a factor for performing a subgroup analysis. However, prokinetic agents were used in some trials [39,40,42,43,46]. So, additional analyses should be performed when sufficient data on bowel preparation regimen can be obtained. Fourth, the start time of initiating education was different among all included trials. This factor and potential influence should be considered with caution. If possible, a further study should be designed to investigate the impact of starting time of education on the detection of colonic polyps or adenomas. Fifth, the EPE methods were quite diverse in the eligible RCTs; however, additional analysis cannot be performed to investigate which EPE method or which combination is the best for bowel preparation instruction because of insufficient data.

Conclusions
In summary, current evidence indicates that there was a significant difference between EPE and standard patient education in PDR (RR 1.19, 95% CI 1.05-1.35; P < .05), ADR (RR 1.37, 95% CI 1.15-1.64; P < .001), ADR-plus (RR 4.39, 95% CI 2.91-6.61; P < .001), and SSADR (OR 1.76, 95% CI 1.22-2.53; P < .05). However, results for AADR (RR 1.54, 95% CI 0.67-3.55; P = .31), ADR-plus, and CDR (RR 1.54, 95% CI 0.60-3.98; P = .37) should be interpreted cautiously as data are still limited. Large-scale trials addressing this question may provide data better applicable to clinical practice.

Acknowledgments
The authors would like to express their gratitude to Ting Shuai and Jing Liang who were invited to critically edit the language of this manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Full search strategy.
[PDF File (Adobe PDF File), 53 KB - mhealth_v8i6e17372_app1.pdf ]

Multimedia Appendix 2
Detailed summary of subgroup analyses.
[PDF File (Adobe PDF File), 905 KB - mhealth_v8i6e17372_app2.pdf ]

Multimedia Appendix 3
Detailed summary of EPE regimes.
[DOCX File , 16 KB - mhealth_v8i6e17372_app3.docx ]

References

https://mhealth.jmir.org/2020/6/e17372

Tian et al

JMIR MHEALTH AND UHEALTH

(footnote text not for citation purposes)


Abbreviations
AADR: advanced adenoma detection rate
ADR: adenoma detection rate
CDR: cancer detection rate
CRC: colorectal cancer
OR: odds ratio
PDR: polyp detection rate
RCT: randomized controlled trial
RR: risk ratio
SSADR: sessile serrated adenoma detection rate
TSA: trial sequential analysis

©Xu Tian, Ling-Li Xu, Xiao-Ling Liu, Wei-Qing Chen. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 01.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Influence of the Business Revenue, Recommendation, and Provider Models on Mobile Health App Adoption: Three-Country Experimental Vignette Study

Abstract

Background: Despite the worldwide growth in mobile health (mHealth) tools and the possible benefits of mHealth for patients and health care providers, scientific research examining factors explaining the adoption level of mHealth tools remains scarce.

Objective: We performed an experimental vignette study to investigate how four factors related to the business model of an mHealth app affect its adoption and users’ willingness to pay: (1) the revenue model (ie, sharing data with third parties vs accepting advertisements); (2) the data protection model (General Data Protection Regulation [GDPR]-compliant data handling vs nonGDPR-compliant data handling); (3) the recommendation model (ie, doctor vs patient recommendation); and (4) the provider model (ie, pharmaceutical vs medical association provider). In addition, health consciousness, health information orientation, and electronic health literacy were explored as intrapersonal predictors of adoption.

Methods: We conducted an experimental study in three countries, Spain (N=800), Germany (N=800), and the Netherlands (N=416), to assess the influence of multiple business models and intrapersonal characteristics on the willingness to pay and intention to download a health app.

Results: The revenue model did not affect willingness to pay or intentions to download the app in all three countries. In the Netherlands, data protection increased willingness to pay for the health app ($P<.001$). Moreover, in all three countries, data protection increased the likelihood of downloading the app ($P<.001$). In Germany ($P=.04$) and the Netherlands ($P=.007$), a doctor recommendation increased both willingness to pay and intention to download the health app. For all three countries, apps manufactured in association with a medical organization were more likely to be downloaded ($P<.001$). Finally, in all three countries, men, younger individuals, those with higher levels of education, and people with a health information orientation were willing to pay more for adoption of the health app and had a higher intention to download the app.

Conclusions: The finding that people want their data protected by legislation but are not willing to pay more for data protection suggests that in the context of mHealth, app privacy protection cannot be leveraged as a selling point. However, people do value a doctor recommendation and apps manufactured by a medical association, which particularly influence their intention to download an mHealth app.

(JMIR Mhealth Uhealth 2020;8(6):e17272) doi:10.2196/17272

Keywords
mHealth adoption; experiment; mobile apps; self-monitoring; privacy paradox; business model; data protection; recommendation; health consciousness; health information orientation; eHealth literacy
**Introduction**

**Background**

Over the last decade, the number of people worldwide who own a mobile phone or another mobile electronic communication device has grown exponentially, fueling the development of mobile health-related services and functions [1,2]. Mobile health (mHealth) [3] can be broadly defined as any medical or public health practice that is supported by mobile devices, ranging from the use of mobile phones to improving points of service data collection, care delivery, and patient communication, to the use of alternative wireless devices for real-time medication monitoring and adherence support (for an overview see [4]). One of the main underlying goals of mHealth is to improve the quality of and access to health care while reducing its costs [5].

Given the potential of mHealth for supporting the health of users, it is important to assess the factors that may motivate or hinder the successful adoption of mHealth technologies and apps. After all, adopting a health technology or app is a first necessary step for ensuring effectiveness [4-6]. However, there is currently insufficient programmatic evidence to inform the implementation and scale-up of mHealth because very little is known about the adoption and effectiveness of mHealth technologies on health [7].

To fill this gap, the aim of this study was to move the field forward by experimentally examining factors that have been suggested to play a role in the adoption of mHealth [8]. We operationalized mHealth adoption in two ways: as having a higher intention to download an mHealth app and being willing to pay a higher price for it. We focus on four factors related to the business model of app development, namely the revenue model, the degree of data protection offered to users, the presence of a doctor recommendation, and whether the app is developed by the pharmaceutical industry or by a medical association. In addition, we explored three intrapersonal characteristics that have been identified as important predictors of electronic health (eHealth) adoption in previous research: health consciousness, health information orientation, and eHealth literacy [9].

Finally, we explored differences among three European countries with varying cultures and health care infrastructures. In Spain, the national health system is an agglomeration of public health services established by the general health law. The vast majority of final providers of care are part of the regional health service structure and are not autonomous legal entities. In Germany, there is a statutory health insurance system that allows people with high incomes to opt out in favor of private coverage. In the Netherlands, there is a statutory health insurance system with universally mandated private insurance (national exchange) that is regulated by the government along with subsidies for insurance. We assume that these differences in national health care infrastructure may impact how users value business models.

**Theoretical Framework**

mHealth can serve multiple purposes such as treatment adherence and disease management, smoking cessation, weight loss, diet, and physical activity [10], thereby providing ample opportunities for people to better monitor and manage their personal health with the aid of their smartphone and other wearable devices [8]. In parallel with the rapid development of mHealth technologies, the focus of health care has shifted from health care providers’ paternalistic approach to a more consumer-oriented approach [11]. At the heart of this approach is the belief that allowing patients to actively access their personal health records and manage their own health will encourage them to be more involved in their own health care [12]. This increased involvement can subsequently strengthen the patient-provider relationship and enhance the (cost-) effectiveness of health care management. Because of these individual and societal benefits associated with mHealth, it is important to gain greater insight into business- and person-level factors that may predict its adoption and use.

**mHealth Business Models**

A first factor related to the business model that may play a role in mHealth adoption is the revenue model. mHealth operates at the intersection of health, technology, and finance, making it a complex industry for the development of sustainable revenue models [5]. Because consumers do not want to spend a large amount of money on the adoption of health apps [13], a great variety of apps have been developed that make revenue on the basis of advertising; however, personal data are also sold to third parties in some cases. Such apps embrace a revenue model that approaches the "privacy as a product" concept [14]. However, it is likely that people experience having their personal health data sold to third parties as a greater “cost” than merely having to accept advertisements in return for “free” access to and use of an mHealth app, as the security of eHealth data is a major concern in the health care industry [5]. Hence, we established the first hypothesis (H1): people are willing to pay more for a health app (H1a) and have a higher intention to download the app (H1b) when they can access and use the app in exchange for accepting advertisements than when having to accept either data sharing with third parties, or a combination of advertising and data sharing with third parties.

In the arena of health care, previous misuses of patient data have affected public confidence in health care research [15]. This was one of the motivating factors for the European Union to implement the General Data Protection Regulation (GDPR) [16]. The GDPR aims to protect people’s right to protection of their data by establishing rules that are related to the free movement of personal data. The GDPR has received widespread public attention in the public domain, and has led to real and significant changes in the ways in which organizations deal with user data. It is reasonable to assume that the GDPR has sharpened citizens’ awareness of and concern for data protection, including when adopting mHealth apps [17]. Hence, we may expect that adoption of a health app will be positively influenced by assurance of adequate protection of personal health data, leading to hypothesis 2 (H2): people are willing to pay more for a health app (H2a) and have a higher intention to download the app (H2b) when the health app ensures data protection in line with European legislation than when no information is given about data protection.
An additional factor that may play a role in the adoption of mHealth apps is whether the app is recommended by medical professionals, who are considered the gatekeepers of health care delivery [18,19]. As an example, in their analysis of factors affecting the adoption of electronic patient records, Raisinghani and Young [20] noted that doctor recommendations were a key factor in the adoption process. Similarly, Peng et al [21] found that patients with type 2 diabetes identified doctor recommendations as a significant factor motivating their adoption of a diabetes mHealth app [22].

There are at least two reasons to explain why a doctor recommendation for a health app can be a strong enforcer for patients to use digital health technologies. First, doctors are considered to be experts in their field of work, and therefore have more influence than nonexperts, particularly since they also know the patients and their interests quite well [19,23]. Second, doctors’ professionalism forces them to act upon the patients’ interests first; most patients therefore trust a doctor more than other actors [24]. Hence, we devised hypothesis 3 (H3): people are willing to pay more for a health app (H3a) and have a higher intention to download the app (H3b) when the app is recommended by doctors than when the app is recommended by a patient association.

Finally, we examined whether a health app manufactured by a medical association is more likely to be adopted than an app manufactured by the pharmaceutical industry. Pharmaceutical companies need to negotiate the conflict between striving for optimal health care and striving for profit [25]. However, in the eyes of the public, it is not always clear that the pharmaceutical industry has patients’ interests at heart [26].

With the advent of mHealth, new concerns have arisen with regard to the quality of these apps, and whether their development and manufacturing should be regulated [27]. With respect to the implementation of mHealth, there are concerns that when the pharmaceutical industry engages in efforts to disseminate health information via mobile devices, they may strategically use these efforts to promote their products and services [28]. In short, given the for-profit nature of the pharmaceutical industry, we may assume that trust in pharmaceutical providers of mHealth apps is generally lower than trust in providers for whom generating profit is not the main goal, such as medical associations or other nonprofit medical associations. This difference in trust may explain a difference in users’ adoption of mHealth apps, leading to hypothesis 4 (H4): people are willing to pay more for a health app (H4a) and have a higher intention to download the app (H4b) when the app is manufactured by a medical association than when the app is manufactured by a pharmaceutical company.

**Personal Factors Affecting mHealth Adoption: Health Consciousness, Health Information Orientation, and eHealth Literacy**

In addition to mHealth business models, we may also consider psychological antecedents that predict adoption [29] to obtain an adequate understanding of personal characteristics that influence the information-use strategies of the online health consumer [30-32]. Studies have shown that the determination to adopt mHealth technologies is greater among people who evaluate their health as more vulnerable to diseases and are more concerned about their health [33], and among people who take more care of their own health [34,35].

According to Dutta-Bergman [34], health consciousness, health information orientation, and eHealth literacy are important factors related to the search for online health information and potentially also to the adoption of a health app. Health consciousness means that an individual takes care of their personal health and that those health concerns are blended into their daily lives [33,36-38]. Health information orientation, defined as the inclination to seek out health information, could be an important predictor to explain who is most willing to adopt a health app [39,40]. Finally, eHealth literacy is considered an important factor predicting health app adoption, since people with higher levels of eHealth literacy have better ability to use health apps [41].

Considering the limited understanding of the general cognitive motivators that trigger people’s usage of health apps, it is important to examine which factors can best explain the adoption of health apps. Therefore, a second aim of this study was to examine whether health consciousness, health information orientation, and eHealth literacy predict the adoption of and willingness to pay for a health app.

**Methods**

**Participants and Design**

We conducted an online vignette experiment in three countries: Spain, Germany, and the Netherlands. Every participant was exposed to four different vignettes, each describing one specific aspect of the business model of an mHealth app (ie, the first with a specific revenue, the second with data protection, the third with a recommendation, and the fourth with a provider model). Next, the likeliness to adopt the health app and willingness to pay were assessed as outcome measures. For each vignette, a different version was randomly assigned to participants. Vignettes describe a hypothetical situation to which participants respond thereby revealing their perceptions, values, attitudes, and intentions. The advantage of vignette studies is a pragmatic and internal valid method assessing participants’ responses to experimental conditions, thereby simulating actual situations the best way possible. Nonetheless, considering that vignettes are a simulation, actual situations might lead to different outcomes. The revenue model was considered between three subject levels (advertising vs data sharing vs advertising and data sharing), data protection was considered between two subject levels (data protection by European Union legislation vs no information), recommendation was considered between two subject levels (recommended by doctors vs patients association), and provider was considered between two subject levels (medical association vs pharmaceutical company). Table 1 shows the distribution of participants over each vignette condition.

---

**Lupiáñez-Villanueva et al**

https://mhealth.jmir.org/2020/6/e17272

JMIIR Mhealth Uhealth 2020 | vol. 8 | iss. 6 | e17272 | p.605

(page number not for citation purposes)
Table 1. Number of participants per condition for Spain, Germany, and the Netherlands.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Spain (N=800)</th>
<th>Germany (N=800)</th>
<th>The Netherlands (N=416)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Business models</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advertising</td>
<td>268</td>
<td>267</td>
<td>146</td>
</tr>
<tr>
<td>Data sharing</td>
<td>265</td>
<td>266</td>
<td>155</td>
</tr>
<tr>
<td>Advertising and data sharing</td>
<td>267</td>
<td>267</td>
<td>115</td>
</tr>
<tr>
<td><strong>Data protection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>European union legislation</td>
<td>400</td>
<td>400</td>
<td>230</td>
</tr>
<tr>
<td>No information provided</td>
<td>400</td>
<td>400</td>
<td>186</td>
</tr>
<tr>
<td><strong>Recommendation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>By doctors</td>
<td>400</td>
<td>400</td>
<td>230</td>
</tr>
<tr>
<td>By patients association</td>
<td>400</td>
<td>400</td>
<td>186</td>
</tr>
<tr>
<td><strong>App provider</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical association</td>
<td>400</td>
<td>400</td>
<td>230</td>
</tr>
<tr>
<td>Pharmaceutical company</td>
<td>400</td>
<td>400</td>
<td>186</td>
</tr>
</tbody>
</table>

The data in Spain (N=800) and Germany (N=800) were collected through an online survey administered by a Spanish professional research company. The sample was chosen through a proportionate stratified sampling method considering gender and age. The data in the Netherlands (N=416) were gathered by snowballing a link of the questionnaire via social media platforms. The participant information is shown in Table 2. Employment status was assessed by the question “Which of these descriptions best describes your situation or applies to what you have been doing for the last month?,” with the answer possibilities ranging from “Employed/Self-employed” to “Another not in the labor force.” We created a bivariate variable with employed vs nonemployed based on this response. Financial status was assessed with the question “During the last 12 months, would you say you had difficulties in paying your bills at the end of the month…?” with the answer possibilities ranging from “Most of the time” to “Never.” All survey participants were informed of the overall study goals and procedures. Only those who agreed to participate in the online survey were given access to the survey. The approval of the Ethical Committee of the university leading the study (Universitat Oberta de Catalunya, Barcelona, Spain) to conduct the experiment was obtained in 2017. We informed participants beforehand that all of the data collected would remain confidential and that they could cease participation at any time.
Table 2. Descriptive information about the participants per country.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Spain (N=800)</th>
<th>Germany (N=800)</th>
<th>The Netherlands (N=416)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (women), n (%)</td>
<td>400 (50.0)</td>
<td>400 (50.0)</td>
<td>159 (38.2)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>41.60 (13.32)</td>
<td>45.92 (15.09)</td>
<td>28.92 (14.86)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary education</td>
<td>26 (3.3)</td>
<td>224 (28.0)</td>
<td>8 (1.9)</td>
</tr>
<tr>
<td>High school diploma</td>
<td>200 (25.0)</td>
<td>285 (35.6)</td>
<td>57 (13.7)</td>
</tr>
<tr>
<td>Some years of university</td>
<td>136 (17.0)</td>
<td>59 (7.4)</td>
<td>105 (25.2)</td>
</tr>
<tr>
<td>University</td>
<td>335 (41.9)</td>
<td>170 (21.3)</td>
<td>175 (42.1)</td>
</tr>
<tr>
<td>Postgraduate degree</td>
<td>103 (12.9)</td>
<td>62 (7.8)</td>
<td>71 (17.1)</td>
</tr>
<tr>
<td>Employment status, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed/self-employed</td>
<td>586 (73.3)</td>
<td>468 (58.5)</td>
<td>161 (38.7)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>64 (8.0)</td>
<td>36 (4.5)</td>
<td>9 (2.2)</td>
</tr>
<tr>
<td>Student</td>
<td>55 (6.9)</td>
<td>55 (6.9)</td>
<td>238 (57.2)</td>
</tr>
<tr>
<td>Retired</td>
<td>60 (7.5)</td>
<td>178 (22.3)</td>
<td>4 (1.0)</td>
</tr>
<tr>
<td>Not working due to illness or disability</td>
<td>12 (1.5)</td>
<td>27 (3.4)</td>
<td>2 (0.5)</td>
</tr>
<tr>
<td>Another not in the labor force</td>
<td>23 (2.9)</td>
<td>36 (4.5)</td>
<td>2 (0.5)</td>
</tr>
<tr>
<td>Financial statusa, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most of the time</td>
<td>123 (15.4)</td>
<td>62 (7.8)</td>
<td>21 (5.0)</td>
</tr>
<tr>
<td>From time to time</td>
<td>293 (36.6)</td>
<td>199 (24.9)</td>
<td>78 (18.8)</td>
</tr>
<tr>
<td>Almost never</td>
<td>375 (46.9)</td>
<td>525 (65.6)</td>
<td>291 (70.0)</td>
</tr>
<tr>
<td>No answer</td>
<td>9 (1.1)</td>
<td>14 (1.8)</td>
<td>26 (6.3)</td>
</tr>
<tr>
<td>Health app use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No use</td>
<td>420 (52.5)</td>
<td>565 (70.6)</td>
<td>285 (68.5)</td>
</tr>
<tr>
<td>One time</td>
<td>105 (13.1)</td>
<td>71 (8.9)</td>
<td>37 (8.9)</td>
</tr>
<tr>
<td>Two times</td>
<td>82 (10.3)</td>
<td>73 (9.1)</td>
<td>28 (6.7)</td>
</tr>
<tr>
<td>Three times</td>
<td>71 (8.9)</td>
<td>36 (4.5)</td>
<td>25 (6.0)</td>
</tr>
<tr>
<td>Four times</td>
<td>33 (4.1)</td>
<td>10 (1.3)</td>
<td>11 (2.6)</td>
</tr>
<tr>
<td>Five times</td>
<td>22 (2.8)</td>
<td>13 (1.6)</td>
<td>4 (1.0)</td>
</tr>
<tr>
<td>More than six times</td>
<td>67 (8.4)</td>
<td>32 (4.0)</td>
<td>26 (6.3)</td>
</tr>
<tr>
<td>Health consciousness, mean (SD)</td>
<td>4.06 (0.78)</td>
<td>3.82 (0.85)</td>
<td>3.43 (0.87)</td>
</tr>
<tr>
<td>Health information orientation, mean (SD)</td>
<td>3.63 (0.83)</td>
<td>3.30 (0.96)</td>
<td>2.79 (1.01)</td>
</tr>
<tr>
<td>eHealthb literate, mean (SD)</td>
<td>3.51 (0.93)</td>
<td>3.50 (0.94)</td>
<td>3.08 (1.03)</td>
</tr>
<tr>
<td>Doctor visits in the last year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>78 (9.8)</td>
<td>95 (11.9)</td>
<td>102 (24.4)</td>
</tr>
<tr>
<td>Once</td>
<td>156 (19.5)</td>
<td>114 (14.2)</td>
<td>87 (20.9)</td>
</tr>
<tr>
<td>Twice</td>
<td>170 (21.3)</td>
<td>151 (18.9)</td>
<td>80 (19.2)</td>
</tr>
<tr>
<td>Three times</td>
<td>128 (16.0)</td>
<td>104 (13.0)</td>
<td>54 (13.0)</td>
</tr>
<tr>
<td>Four times</td>
<td>86 (10.8)</td>
<td>97 (12.1)</td>
<td>31 (7.5)</td>
</tr>
<tr>
<td>Five times or more</td>
<td>182 (22.8)</td>
<td>236 (29.5)</td>
<td>57 (13.7)</td>
</tr>
<tr>
<td>General health, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very bad</td>
<td>1 (0.1)</td>
<td>11 (1.4)</td>
<td>143 (34.4)</td>
</tr>
<tr>
<td>Bad</td>
<td>21 (2.6)</td>
<td>78 (9.8)</td>
<td>223 (53.6)</td>
</tr>
</tbody>
</table>
Health consciousness was measured using 5 statements that were each rated on a 5-point scale (1 strongly disagree to 5 strongly agree) [42]. Reliability of the scale was high (Cronbach \(\alpha=.95\)).

**Statistical Analyses**

Multiple linear regression models were conducted for every country and the two dependent variables (willingness to pay and intention to download) separately. Within each regression, in the first step we assessed the effect of the business model; in the second step we included age, gender, and education; and in the third step we included health consciousness, health information orientation, and eHealth literacy. In addition, effect sizes were calculated for each regression model. Following Cohen [43], an effect size of \(R^2\) around 0.1 is interpreted as low, \(R^2\) varying around 0.3 is considered medium, and \(R^2>0.5\) is interpreted as a large effect.

**Results**

**Revenue Models**

Linear regression analyses were first conducted to explore the role of the revenue model for Spain (see Multimedia Appendix 1). We first examined if Spanish people were more willing to pay more for a health app when they could access and use the app in exchange for accepting advertisements than when having to accept data sharing with third parties, either individually or with a combination of advertising (H1a). No effects were found for data sharing (\(P=.20\)) or data sharing and advertising (\(P=.50\)) as revenue models (advertising was used as the reference category) on willingness to pay. Furthermore, men (\(P=.02\)) and people with a health information orientation (\(P=.002\)) were more willing to pay more for the health app. The explained variance for the model including all predictors was 22.4%. These findings do not support H1a and H1b.

Similar results were obtained in the analyses for the German sample (see Multimedia Appendix 2). No effects were found for data sharing (\(P=.23\)) or data sharing and advertising (\(P=.07\)) as revenue models (advertising as the reference category) on willingness to pay (H1a) in all three models. Furthermore, younger people (\(P=.02\)), people who obtained a postgraduate degree (\(P=.01\)) compared to students, those who were employed
(P=.02), and people with a health information orientation (P<.001) were more willing to pay more for adopting the health app. The explained variance for the model including all predictors was 7.94%. Next, we tested the same model but with intention to download the app as the outcome measure (H1b). Again, no effects were found for data sharing (P=.26) or data sharing and advertising (P=.08) as revenue models on the intention to download in all three models. Men (P=.02), younger people (P<.001), people who finished high school (P=.01) or university (P=.003), those who were employed (P<.001), and people with a health information orientation (P<.001) reported greater intentions to download the health app. The explained variance for the model including all predictors was 29.7%. These findings do not support H1a and H1b.

Similar results were also obtained for the Netherlands (see Multimedia Appendix 3). No effects were found for data sharing (P=.38) or data sharing and advertising (P=.17) as revenue models (advertising as the reference category) on willingness to pay in all models. Furthermore, men (P=.02), those who were employed (P=.01), and people with a health information orientation (P=.05) were willing to pay more for adopting the health app. The explained variance for the model including all predictors was 5.3%. For intentions to download the app, no effects were found for data sharing (P=.38) or data sharing and advertising (P=.43) as revenue models in all three models. People with a health information orientation had greater intentions to download the health app (P=.05). The explained variance for the model including all predictors was 3.5%. These findings do not support H1a and H1b.

Data Protection Models

Next, we explored the role of the data protection model. Linear regression analyses were first performed for Spain (see Multimedia Appendix 4) to examine if people were willing to pay more for a health app (H2a) and had a higher intention to download the app (H2b) when the health app ensured data protection in line with European legislation than when no information was given about data protection. No effects were found for the data protection model (no information about data protection was the reference category) on willingness to pay in all three models. Furthermore, men (P=.03) and people with a health information orientation (P=.006) were willing to pay more for adopting the health app. The explained variance for the model including all predictors was 3.0%. In contrast, participants in the condition whereby data were protected by European Union legislation had greater intentions to download the app (P<.001) in all three models. Furthermore, men (P=.02), younger people (P<.001), people who finished high school (P=.01), university (P<.001), or had a postgraduate degree (P=.03); and people with a health information orientation (P<.001) reported greater intentions to download the health app. The explained variance for the model including all predictors was 33.6%. These results do not support H2a but do support H2b.

Finally, linear regression analyses were conducted for the Netherlands (see Multimedia Appendix 6). Participants in the condition whereby data protection by European Union legislation was explicitly stated were more willing to pay more for the health app than participants who received no information about data protection (P<.001). No significant effects were found for the other factors. The explained variance for the model including all predictors was 5.7%. In addition, participants in the condition where data were protected by European Union legislation reported greater intentions to download (P<.001) in all three models. Furthermore, people with a health information orientation had a greater intention to download the health app (P=.003). The explained variance for the model including all predictors was 11.2%. These results support both H2a and H2b.

Recommendation Models

Next, we explored the role of the recommendation model. The first linear regression analyses were conducted for Spain (see Multimedia Appendix 7) to examine if people were more willing to pay more for a health app (H3a) and had a higher intention to download the app (H3b) when the app was recommended by doctors than when the app was recommended by a patient association (reference category). No effects were found for the recommendation model on willingness to pay in all three models. Furthermore, men (P=.02) and people with a health information orientation (P=.01) were willing to pay more for adopting the health app. The explained variance for the model including all predictors was 3.4%. In contrast, participants reported greater intentions to download the health app when doctors recommended the health app (P=.04) compared to when the patients association recommended the health app. Furthermore, men (P=.02), younger people (P<.001), those who were employed (P=.01), and people with a health information orientation (P<.001) reported greater intentions to download the health app. The explained variance for the model including all predictors was 21.2%. These results do not support H3a but do support H3b.

Next, we conducted linear regression analyses for Germany (see Multimedia Appendix 8). Participants were more willing to pay for the app when it was recommended by doctors than when it was recommended by the patients association (P=.02), except in the model including all predictors. Furthermore, people who finished university (P=.007) compared to students, and people with a health information orientation (P<.001) were willing to pay more for adopting the health app. In contrast, people with more eHealth literacy were less willing to pay more for the health app (P=.007). The explained variance for the model including all predictors was 7.6%. Participants had greater
intentions to download the health app when it was recommended by doctors compared to when it was recommended by the patients' association ($P=.01$), but only when not controlling for sociodemographic and dispositional factors. Furthermore, men ($P=.02$); younger people ($P<.001$); people who finished high school ($P=.001$) and university ($P<.001$), or those with a postgraduate degree ($P=.009$) compared to students; those who were employed ($P<.001$); and people with a health information orientation had greater intention to download the health app ($P<.001$). The explained variance for the model including all predictors was 31.2%. These results support both H3a and H3b.

Finally, linear regression analyses were conducted for the Netherlands (see Multimedia Appendix 9). Participants were willing to pay more for the health app when it was recommended by doctors compared to when it was recommended by the patients' association ($P=.01$) in all three models. No significant effect was found for the other factors. The explained variance for the model including all predictors was 2.7%. In addition, participants reported greater intentions to download the health app when it was recommended by doctors compared to when it was recommended by the patients' association ($P=.01$). Furthermore, people with a health information orientation had greater intentions to download the health app ($P<.001$). The explained variance for the model including all predictors was 8.6%. These results support both H3a and H3b.

### Provider Models

Finally, we conducted linear regression analyses to explore the role of the provider model. The first set of analyses were conducted for Spain (see Multimedia Appendix 10). We examined if people were more willing to pay more for a health app (H4a) and had a higher intention to download the app (H4b) when the app was manufactured by a medical association than when it was manufactured by a pharmaceutical company (reference category). No effects were found for the provider model on willingness to pay in all three models. Furthermore, men ($P=.03$) and people with a health information orientation ($P=.004$) were willing to pay more for adopting the health app. The explained variance for the model including all predictors was 3.5%. In contrast, participants had less intention to download the health app if it was provided by a medical association compared to when it was provided by a pharmaceutical company ($P<.001$) in all three models. Furthermore, men ($P=.06$), younger people ($P=.001$), those who were employed ($P=.009$), and people with a health information orientation ($P<.001$) had greater intentions to download the health app. The explained variance for the model including all predictors was 23.1%. These results do not support H4a but do support H4b.

The second set of linear regression analyses were conducted for Germany (see Multimedia Appendix 11). No significant effects were found for the provider model on willingness to pay in all three models. Furthermore, younger people ($P=.005$), people with a health information orientation ($P<.001$), and people with less eHealth literacy ($P=.002$) were willing to pay more for adopting the health app. The explained variance for the model including all predictors was 6.3%. In contrast, participants reported less intentions to download the health app when it was provided by a medical association compared to when it was provided by a pharmaceutical company ($P<.001$) in all three models. Furthermore, men ($P=.002$); younger people ($P<.001$); people who finished high school ($P=.005$), university ($P<.001$), or had a postgraduate degree ($P=.03$), with students as reference; people with a health information orientation ($P<.001$); and people with less eHealth literacy ($P=.02$) had greater intentions to download the health app. The explained variance for the model including all predictors was 29.9%. These results do not support H4a but do support H4b.

Finally, linear regression analyses were conducted for the Netherlands (see Multimedia Appendix 12). Participants were willing to pay more for the health app when it was provided by a medical association compared to when it was provided by a pharmaceutical company ($P=.005$) in all three models. No significant effects were found for the other factors. The explained variance for the model including all predictors was 5.0%. In addition, participants reported greater intentions to download the app when it was provided by a medical association than when it was provided by a pharmaceutical company ($P<.001$) in all three models. Furthermore, people with a health information orientation had greater intention to download the health app ($P=.008$). The explained variance for the model including all predictors was 7.0%. These results support both H4a and H4b.

### Discussion

#### Principal Findings

Given the expected benefits associated with mHealth adoption, both for individual users and health care systems, it is important to gain greater understanding of factors that contribute to or deter from adoption. Therefore, we conducted an online experiment to assess the effect of four variations in the business model of an mHealth app and three intrapersonal characteristics in three different countries (Spain, Germany, and the Netherlands) on individuals' willingness to pay for and their likelihood of adopting an mHealth app.

The results showed that in all countries there was no effect of the different revenue models on both willingness to pay and intention to download the health app, thereby not supporting H1. People are not less willing to pay and do not have a reduced intention to download a health app when the revenue model is based on data sharing or advertising and data sharing, compared to that based on advertising only. This finding is surprising, as people in general report being concerned about sharing their personal information [42]. Hence, this concern could be expected to drive one's intended and actual disclosure, and their subsequent decision making. Our study does not support this speculation, and instead suggests that people are less than selective and often cavalier in the protection of their own data profiles. To date, few studies have examined this discrepancy between individuals' intentions to protect their own privacy and how they actually behave in the marketplace, which is termed the "privacy paradox" (see [42]) in the context of mHealth. Our findings indicate that further research on this matter is warranted, given that the privacy paradox is an increasing concern when it comes to personal health data [11,44].
Interestingly, in Spain and Germany, we found no effects of the data protection model on willingness to pay, whereas in the Netherlands, participants in the data protection condition were willing to pay more for the health app compared to when receiving no information regarding how their health information will be used. In all three countries, participants in the data protection condition were more likely to download the health app, thereby partly supporting H2. Thus, in line with the findings for the revenue model, and supporting the notion of the privacy paradox, people were not willing to pay for the app. Given an industry in which mobile apps are continuously expanding and new health care apps and devices are rapidly being created, it is essential to be very cautious of the collection and treatment of users’ personal health information, particularly by the consumers themselves [44].

In Spain, we found no effect of the recommendation model, whereas in Germany and the Netherlands, participants were willing to pay more for a health app recommended by a doctor compared to that recommended by a patient association. In addition, in all three countries, intentions to download the health app were greater when the app was recommended by a doctor.

In Spain and Germany, we found no effects of the provider model, whereas in the Netherlands, participants in the medical association provider condition were willing to pay more for the health app than participants in the pharmaceutical provider condition. In all three countries, participants in the medical association provider condition had greater intentions to download the app than participants in the pharmaceutical provider condition.

Overall, the findings of our study indicate that endorsement from the medical establishment, either via a doctor recommendation or a medical association provider model, is helpful to increase adoption of an mHealth app. However, the revenue and data protection models seem to have a less consistent and a weaker effect, especially on the willingness to pay for an app. These findings suggest that future app developers can benefit most from a close collaboration with medical experts and organizations to increase adoption rates.

In general, the above findings show that certain aspects of the business model can influence the willingness to pay for or the intention to adopt an mHealth app, but that this influence appears conditional; that is, it varies according to the country of residence and seems to interact with dispositional characteristics such as a person’s health information orientation. In summary, this suggests that mHealth adoption is a complex process that involves many different factors situated at least at the personal, economic, and cultural level. This implies that in order to increase adoption rates and decrease attrition, developers, organizations, and practitioners need to be weary of one-size-fits-all approaches, as these are likely less successful than an approach that tailors the business model to the population of interest. Given that we currently lack understanding of the precise mechanisms that explain why, under certain conditions, mHealth adoption can be more or less successful, future research is needed to explore these mechanisms in greater depth.

Finally, in all three countries, men, younger individuals, people with higher levels of education, and those with a health information orientation were willing to pay more for adoption of the health app and had a higher intention to download the app. In line with previous studies, health information orientation was found to be an important predictor that explains both the willingness to pay and the intention to download the health app [36,39,40]. A high level of health information orientation positively affected the amount a participant was willing to pay for the health app and the intention to download it.

Overall, the finding that young, highly educated males, and people with a stronger health information orientation were more willing to pay for and download the mHealth app in this study suggests that traditional factors that demarcate access to and use of health services such as gender and age are also at play in mHealth. Owing to the ease of use and widespread diffusion of mobile phones, mHealth initiatives are often applauded for their emancipatory potential (eg, [45]); thus, our study supports earlier observations that future policy efforts aimed at closing “the digital health divide” need to also focus on disparities in mHealth adoption and use [46]. To inform these policy efforts, further research is needed to explore the specific barriers hindering participation in mHealth.

Strengths and Limitations

One of the strengths of the current study is that we collected data among a large group of participants in three different countries. Another strength is that we used a multifactorial experimental design, examining several factors in relation to the business model that are considered to be important in predicting and explaining the adoption of a health app. Third, we assessed the role of three intrapersonal predictors of the adoption of a health app.

This study also had some limitations. First, because the study was conducted online, internal validity of the exact experiment cannot be guaranteed since it is difficult to assess how truthfully the participants answered. Nonetheless, because the experiment was not focused on sensitive questions but rather on factors related to adoption of an online health app, using an online questionnaire to assess different factors could be considered a valid and reliable measurement. Second, both in Spain and Germany, data were collected by a professional company in which the participants were paid for their participation, whereas in the Netherlands a convenience sampling approach was used without paying the participants. Overall, the results are quite similar between the countries, although we also noticed some minor differences in some results that could be due to the different sampling methods.

Conclusion

Over the last decade, the number of people in the world who perform health-related functions on their smartphones has increased rapidly [1,2]. However, research into the adoption and effectiveness of mHealth remains scarce. This is unfortunate, given that adopting a health app is a necessary first step for such an app to be effective [6,44,47]. It is essential that patient safety (data protection), reducing costs, and creating sound business models are investigated to a larger extent to gain

https://mhealth.jmir.org/2020/6/e17272
a better understanding of the major driving forces for the adoption of mHealth in the future. Next, it is important to create standards for mobile apps, whereby doctors and patients associations can have a leading role in informing the potential consumers as a heuristic approach. Governments, large funders, and industry associations should create and adhere to such standards so that mHealth apps can be adopted and used with confidence of the quality, privacy of the data, and with prices that are proportional to the service provided.

Acknowledgments
The current study is part of a broader project titled “mHealth: Challenges and opportunities for health systems.” This work was supported by RecerCaixa of “La Caixa” Foundation.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Linear regression analyses with willingness to pay and intention to download for the business models in Spain.
[DOCX File, 14 KB - mhealth_v8i6e17272_app1.docx]

Multimedia Appendix 2
Linear regression analyses with willingness to pay and intention to download for the business models in Germany.
[DOCX File, 14 KB - mhealth_v8i6e17272_app2.docx]

Multimedia Appendix 3
Linear regression analyses with willingness to pay and intention to download for the business models in the Netherlands.
[DOCX File, 14 KB - mhealth_v8i6e17272_app3.docx]

Multimedia Appendix 4
Linear regression analyses with willingness to pay and intention to download for the data collection models in Spain.
[DOCX File, 14 KB - mhealth_v8i6e17272_app4.docx]

Multimedia Appendix 5
Linear regression analyses with willingness to pay and intention to download for the data collection models in Germany.
[DOCX File, 14 KB - mhealth_v8i6e17272_app5.docx]

Multimedia Appendix 6
Linear regression analyses with willingness to pay and intention to download for the data collection models in the Netherlands.
[DOCX File, 14 KB - mhealth_v8i6e17272_app6.docx]

Multimedia Appendix 7
Linear regression analyses with willingness to pay and intention to download for the recommendation models in Spain.
[DOCX File, 14 KB - mhealth_v8i6e17272_app7.docx]

Multimedia Appendix 8
Linear regression analyses with willingness to pay and intention to download for the recommendation models in Germany.
[DOCX File, 14 KB - mhealth_v8i6e17272_app8.docx]

Multimedia Appendix 9
Linear regression analyses with willingness to pay and intention to download for the recommendation models in the Netherlands.
[DOCX File, 14 KB - mhealth_v8i6e17272_app9.docx]

Multimedia Appendix 10
Linear regression analyses with willingness to pay and intention to download for the provider models in Spain.
[DOCX File, 14 KB - mhealth_v8i6e17272_app10.docx]

Multimedia Appendix 11
Linear regression analyses with willingness to pay and intention to download for the provider models in Germany.

[DOCX File, 14 KB - mhealth_v8i6e17272_app11.docx]

Multimedia Appendix 12
Linear regression analyses with willingness to pay and intention to download for the provider models in the Netherlands.

[DOCX File, 14 KB - mhealth_v8i6e17272_app12.docx]

References


Abbreviations

eHealth: electronic health
GDPR: General Data Protection Regulation
H1: hypothesis 1
H2: hypothesis 2
H3: hypothesis 3
mHealth: mobile health

©Francisco Lupiáñez-Villanueva, Frans Folkvord, Mariek Vanden Abeele. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 04.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Feasibility of a Home-Based Tablet App for Dexterity Training in Multiple Sclerosis: Usability Study

Judith Jantine Willemijn van Beek, MSc; Erwin Everardus Henri van Wegen, PhD; Marc Berend Rietberg, PhD; Thomas Nyffeler, MD; Stephan Bohlhalter, MD; Christian Philipp Kamm, MD; Tobias Nef, PhD; Tim Vanbellingen, PhD

Neurocenter, Lucerne Cantonal Hospital, Lucerne, Switzerland
Gerontechnology and Rehabilitation Group, University of Bern, Bern, Switzerland
Department of Rehabilitation Medicine, Amsterdam Movement Sciences, Multiple Sclerosis Center Amsterdam, Amsterdam University Medical Centers, Vrije Universiteit Medical Center, Amsterdam, Netherlands
Department of Neurology, Bern University Hospital, Bern, Switzerland

Corresponding Author:
Tim Vanbellingen, PhD
Neurocenter
Lucerne Cantonal Hospital
Spitalstrasse
Lucerne, 6000 16 Lucerne
Switzerland
Phone: 41 41 205 56 34
Email: tim.vanbellingen@luks.ch

Abstract

Background: Persons with multiple sclerosis (MS) often experience dexterous difficulties during the performance of activities of daily living, such as fastening buttons, handling coins, or writing, therefore impacting their health-related quality of life. Mobile health (mHealth) solutions, such as tablet apps, may be used to train impaired dexterous skills. The feasibility of a tablet app–based dexterity home-based intervention in MS (TAD-MS) has not been explored yet in persons with MS.

Objective: The aim of this pilot study was to evaluate the feasibility and usability of home-based dexterity training with a tablet app in both persons with MS and healthy subjects.

Methods: A total of 9 persons with MS, aged 35-71 years, with an Expanded Disability Status Scale score between 2 and 7.5, performed the TAD-MS for 4 weeks, five times a week, with each training session lasting approximately 30 minutes. Participants’ impaired dexterity was measured by the Nine-Hole Peg Test. A total of 10 age-matched healthy subjects also tested and rated the usability of the app. Outcome measures were the adherence rate as well as usability measured by the System Usability Scale and a Custom User Engagement Questionnaire (CUEQ).

Results: High feasibility of the tablet app–based dexterity training program was shown by a 97% adherence rate to the training protocol (ie, mean 19.4/20 sessions completed, SD 0.8). High system usability scores (ie, mean 85.39%, SD 11.67) and overall high scores given in the CUEQ (ie, mean 8.2/10, SD 1.4) further point to high usability of the app. Neither demographic variables nor dexterity levels affected the use of the app.

Conclusions: This pilot study is the first to demonstrate high feasibility and usability of a new tablet app–based dexterity home-based training program among both persons with MS and healthy individuals. Whether this kind of training improves dexterity will need to be evaluated in a randomized controlled trial.

(JMIR Mhealth Uhealth 2020;8(6):e18204) doi:10.2196/18204

KEYWORDS
dexterity; feasibility; multiple sclerosis; rehabilitation; app; home-based training
**Introduction**

Multiple sclerosis (MS) is the most common cause of nontraumatic neurological disability in young adults and comes with long-term disability, including dexterous dysfunction [1-3]. Impaired dexterity is a frequently reported problem in MS, affecting approximately 70%-80% of persons with MS [4], mainly due to deconditioning, coordination deficits, muscle weakness, and/or tremor [5]. Subsequently, persons with MS often experience problems in performing activities of daily living (ADL), such as fastening buttons, handling coins, or writing, therefore negatively impacting their self-reported health-related quality of life [6].

In the last decade, the implementation of technical innovations has rapidly grown within MS neurorehabilitation. Many of these innovations relate to mobile health (mHealth) solutions and are often apps that can easily be installed on smartphones and/or tablets [7] for use in the home environment. Their use is increasing among today’s MS population [7]. Recently, Marrie and colleagues showed in a survey, which included over 6000 persons with MS, that around 84% already used mHealth services for exchanging medical information [8]. Approximately half of the smartphone and tablet users used an mHealth app. Interestingly, a large majority of users (98.7%) reported some kind of benefit (eg, increased physical activity or being better informed about diagnosis) [8].

The most frequently included features in such apps focus on personal data management and patient education (ie, health status) by providing, for example, diaries [9,10]. Some apps directly tackle physical or cognitive performance, such as the Cognitive Training Kit (COGNI-TRAcK) app [11]. This app was developed to train cognitive deficits, which are common among persons with MS [12,13]. COGNI-TRAcK showed high feasibility and usability as well as improved working memory if the app was used within a home-based cognitive training environment [14]. The home-based training environment certainly added to the feasibility, because the persons with MS were flexible in planning their sessions and saved on travel time to rehabilitation centers. The specific contents of the app were also experienced to mentally stimulate the users, which helped to sustain motivation, and the contents were fun to play, explaining the high adherence to the training protocol [14].

With regard to physical performance, Thirumalai and colleagues recently developed the therapeutic exercise TEAMS (Tele-Exercise and Multiple Sclerosis) app, aiming to be a comprehensive home tele-exercise program [15]. It contains several exercise videos and articles, explaining yoga, Pilates, and dual-tasking [15]. The advantage of this app is that the physical activity of persons with MS at home can be enhanced through tele-exercising. However, the primary aim of this app was to promote overall physical activity and not dexterity. So far, one app to measure dexterity in MS [16] and two dexterity training apps [17,18] have been developed and tested for their feasibility, either among stroke survivors [17] or among people after surgical carpal tunnel release [18]. The content of these training apps was primarily based on tapping or pinching performance of single digits [17,18]. An advantage of both apps was that finger movements, performed while touching the touchscreen, could be guided by feedback. To our knowledge, the usability of a training dexterity app has not been explored yet in persons with MS.

This pilot study comprehensively evaluated the feasibility and usability of a new tablet app–based dexterity home-based intervention in MS (TAD-MS) [6] in both persons with MS as well as healthy subjects. The exercises focus on various components of dexterity, such as the pinch grip, coordinated finger movements, and tapping movements, and are designed to be challenging and fun. We hypothesized that the new dexterity app is feasible and usable, for both persons with MS and healthy individuals. Furthermore, we expected high adherence to the home-based training protocol.

**Methods**

**Participants and Setting**

For this pilot study, both persons with MS and healthy subjects participated. Patients were prospectively recruited from the Neurocenter, Lucerne Cantonal Hospital, Lucerne, Switzerland. They were included if they met the following inclusion criteria: (1) provided written informed consent according to the latest Declaration of Helsinki [19], (2) had an MS diagnosis following the revised McDonald criteria (ie, primary or secondary progressive or relapsing-remitting MS) [20], (3) experienced impaired dexterity in their ADL or had a pathological Nine-Hole Peg Test (9HPT) according to established cutoff values [21], (4) aged above 18 years, and (5) were able to understand German for instructions and measurements. Patients were excluded if they (1) suffered from severe cognitive impairment (ie, Montreal Cognitive Assessment score less than 21 [22]) or (2) had other neurological, psychiatric, or developmental diseases prior to MS diagnosis. The Ethics Committee of Northwest and Central Switzerland (EKNZ) approved this study. We aimed to recruit age- and gender-matched healthy subjects, who worked as health care professionals or other staff members at the cantonal hospital, to explore whether usability scores of the TAD-MS differed between persons with MS and healthy individuals.

**Tablet App–Based Dexterity Training**

For eligible participants, an intake meeting in our clinic was planned. Upon consent, descriptive characteristics were assessed and participants were asked whether they were having other treatment sessions (ie, physiotherapy or occupational therapy). If yes, all subjects were allowed to participate if their regime was not related to dexterity, since that would have interfered with the outcome of this study. Subsequently, all patients received instructions about the TAD-MS by an expert clinician (JJWvB). The intervention consisted of five 30-minute sessions per week for 4 weeks; therefore, in total, 20 home sessions were completed. Patients received a tablet with the installed app and a personalized log-in code. An Apple iPad (Manufacturer Part Number: MP2G2TY; 9.7-inch Retina display; Wi-Fi; 2017) or Samsung Galaxy Tab S2 (model: SM-T813; 9.7-inch display; Wi-Fi) was used. The app was developed by a team of experts, consisting of an occupational therapist (Amy Orellana), see Acknowledgments, two graphic designers, and one
programmer. In line with the Technology Acceptance Model, which is a frequently used acceptance framework for understanding clinicians’ adoption of mHealth [23], perceived usefulness, perceived ease of use, and enjoyment were key elements that were integrated into this app. The programmer wrote the software program for the app with the programming language C# Job System using Unity 5.0 (Unity Technologies) [24]. The team met on a weekly basis to test and retest the function of the exercises, to resolve bugs, and to review the progress of the software.

Besides an explanation of the exercises from the clinician (JJWvB), the participant got instructions through instructional videos and texts for each exercise separately in the app, explaining how to perform the exercise. An example of the exercises is shown in Figure 1. For a detailed description of the exercises, we refer to our protocol paper where we comprehensively explained the six dexterous games [6]. The clinician (JJWvB) checked the adherence to the protocol by logging in to a website where the performance of the participants was tracked and recorded.

The healthy subjects were invited to our clinic, and after receiving instructions by a trained clinician (Cleo Bol, see Acknowledgments), they performed all the dexterous exercises in a single session. Both patients and healthy subjects did all the exercises with both hands separately, starting with their dominant hand. Patients could choose for themselves when to do the exercises (ie, morning, afternoon, or evening) but were instructed to not train twice on the same day.

**Figure 1.** The TAD-MS app contains six dexterity exercises, including selective finger tapping (balloons), which is shown here. Other exercises are acrobat seesaw (pinch grip), the wheel (rotation and flexion/extension in various joints in fingers), and elevator I (selective finger rotation). TAD-MS: tablet app–based dexterity home-based intervention in multiple sclerosis.

**Objectives**

Our first objective was to extensively evaluate the feasibility and usability of a specific tablet app–based dexterity home-based training program in both persons with MS and healthy subjects. Our second objective was to assess the adherence to the protocol. Our third objective was to investigate whether demographic variables (ie, age and disease duration, type, and severity) and dexterous abilities could interfere with the use of the app.

**Endpoints**

All measurements were collected and rated by a single expert clinician (JJWvB) to achieve an optimal standardization of outcome measurements. The endpoints were feasibility measures: (1) the adherence to the protocol, (2) the self-reported System Usability Scale (SUS) score, and (3) a Custom User Engagement Questionnaire (CUEQ) (see Multimedia Appendix 1), where the opinions of the participants were sought, regarding the content of the app exercises as well as the home-based training.

Based on our previously published pilot study [25], the adherence rate was determined as the ratio of the number of home sessions performed (SP) and the prescribed number of sessions (PS), which is 20: SP/PS × 100% = SP/20 × 100%. An adherence rate of 80% or higher was considered as good.

The SUS was filled out after completing the intervention. The SUS is a well-validated questionnaire, consisting of a 10-item,
Results

Recruitment took place between April 2018 and May 2019. A total of 9 persons with MS started and completed the intervention. No adverse effects or complications and no dropouts were reported. In addition, 10 age-matched healthy subjects participated (mean age 52 years, SD 13, range 33-73). Detailed clinical and demographic characteristics of the persons with MS are presented in Table 1.

The adherence to the training protocol by the persons with MS with regard to the average total sessions completed was very high. More specifically, the persons with MS completed, on average, 19.4 sessions (SD 0.8) of the planned 20 sessions (97%).

Regarding the self-reported system usability of the app, the mean SUS scores were 85.00% (SD 11.59) and 85.75% (SD 12.36) for the patients and the healthy subjects, respectively, indicating excellent usability (see Figure 2). There was no significant difference between these scores (Mann-Whitney test, P=.74). Out of 10 healthy subjects, 2 (20%) rated the system just below the cutoff of 71.4%, meaning acceptable usability. Out of 9 patients, 1 (11%) rated the app as 60%, which means moderate usability. The overall SUS score (N=19) was 85.39% (SD 11.67).

The CUEQ revealed high scores on each dexterity-related question, and the overall quality of the app was scored with a mean of 8.20 (SD 1.35), further indicating very good usability. Only 1 person out of 9 (11%) with MS gave it a score of 5, indicating moderate usability. The others (8/9, 89%) gave it a score of 7 or higher. None of the participants needed additional technical support at home, further denoting the usability. The mean quality scores of the dexterity-related questions for both groups are shown in Figure 3.

We did not find significant correlations between the SUS scores and age (r=.20, P=.42), gender (eta=.024), disease type (eta=.16), disease severity (r=.46, P=.22), disease duration (r=.34, P=.38), AMSQ score (r=.59, P=.09), and 9HPT time (r=.59, P=.10).

Statistical Analyses

Descriptive statistics were used to present the baseline and clinical characteristics of all participants. The observed data approach was used for missing data. The normality of distribution of all outcome measurements was checked by Shapiro-Wilk tests. Moreover, skewness and kurtosis of the residuals were used to further check for normal distribution of outcome scores. To explore relationships between clinical parameters (ie, age, gender, disease type, disease severity, disease duration, AMSQ score, and 9HPT time) and usability scores, Spearman or eta correlations were performed.

The level of significance was set at P≤.05 (two-tailed). Statistical analyses were performed using SPSS Statistics for Windows, version 26.0 (IBM Corp.).

5-point Likert scale—each item was scored from 0 (strongly disagree) to 4 (strongly agree)—which takes three usability criteria into account: effectiveness, satisfaction, and efficiency [26,27]. The total score is obtained by multiplying the mean sum value by 2.5. The SUS score ranges from 0% to 100%, where a higher score indicates better system usability. A score of 70% up to a maximum of 100% represents acceptable-to-excellent usability [26,27].

The CUEQ contained two parts. Firstly, seven specific questions (Qs) related to the content of the app had to be scored on a scale ranging from 1 (poor) to 5 (excellent) (ie, Q1 = Were the exercises fun to play?; Q2 = Have you improved your fine motor skills?; Q3 = Was it easy to integrate the exercises into daily life?; Q4 = Were the explanations for the execution of each exercise sufficient?; Q5 = Would you recommend this application?; Q6 = Can you take the tablet well and easily on the go?; and Q7 = Do you notice improvements in everyday life regarding fine motor skills?). Since the healthy subjects did not participate in the training protocol, they did not answer Q2 and Q7. Secondly, all participants rated the overall quality of the app, which ranged from 1 (very poor quality) to 10 (excellent quality). Furthermore, they were asked to make suggestions for further improvement regarding the content of the app. All questions of the CUEQ were defined based on agreement between two experts in the field of dexterity rehabilitation (TV and Amy Orellana, see Acknowledgements); therefore, they were based on face validity and content validity.

Besides demographics, dexterous functions were measured in the patient cohort to get an impression of overall disability and to check whether impaired dexterity may interfere with the use of the app. We used a patient-reported outcome measure (ie, the Arm Function in Multiple Sclerosis Questionnaire [AMSQ]), a performance-based task (ie, the 9HPT), and a measure of the degree of neurologic impairment (ie, the Expanded Disability Status Score [EDSS]). The 9HPT [28] and the AMSQ [29,30] have shown good reliability and validity in MS. The AMSQ consists of 31 questions on a unidimensional scale that are formulated as “During the past two weeks, to what extent has MS limited your ability to...?” with a 6-point response option ranging from 1 (not at all limited) to 6 (no longer able to). One final score is obtained, ranging from 31 to 186, with a higher score indicating more difficulties in dexterous function [29,30].

For the 9HPT, the time to perform the task was measured in seconds for each hand separately. Normative values considering age and gender are available [21]. The EDSS examines levels of disability [31]. (Scores range from 0 (normal neurological examination) to 10 (death due to MS) in 0.5-unit increments that represent higher levels of disability [31].

The CUEQ contained two parts. Firstly, seven specific questions (Qs) related to the content of the app had to be scored on a scale ranging from 1 (poor) to 5 (excellent) (ie, Q1 = Were the exercises fun to play?; Q2 = Have you improved your fine motor skills?; Q3 = Was it easy to integrate the exercises into daily life?; Q4 = Were the explanations for the execution of each exercise sufficient?; Q5 = Would you recommend this application?; Q6 = Can you take the tablet well and easily on the go?; and Q7 = Do you notice improvements in everyday life regarding fine motor skills?). Since the healthy subjects did not participate in the training protocol, they did not answer Q2 and Q7. Secondly, all participants rated the overall quality of the app, which ranged from 1 (very poor quality) to 10 (excellent quality). Furthermore, they were asked to make suggestions for further improvement regarding the content of the app. All questions of the CUEQ were defined based on agreement between two experts in the field of dexterity rehabilitation (TV and Amy Orellana, see Acknowledgements); therefore, they were based on face validity and content validity.

Besides demographics, dexterous functions were measured in the patient cohort to get an impression of overall disability and to check whether impaired dexterity may interfere with the use of the app. We used a patient-reported outcome measure (ie, the Arm Function in Multiple Sclerosis Questionnaire [AMSQ]), a performance-based task (ie, the 9HPT), and a measure of the degree of neurologic impairment (ie, the Expanded Disability Status Score [EDSS]). The 9HPT [28] and the AMSQ [29,30] have shown good reliability and validity in MS. The AMSQ consists of 31 questions on a unidimensional scale that are formulated as “During the past two weeks, to what extent has MS limited your ability to...?” with a 6-point response option ranging from 1 (not at all limited) to 6 (no longer able to). One final score is obtained, ranging from 31 to 186, with a higher score indicating more difficulties in dexterous function [29,30].

For the 9HPT, the time to perform the task was measured in seconds for each hand separately. Normative values considering age and gender are available [21]. The EDSS examines levels of disability [31].
Table 1. Clinical and demographic characteristics of the persons with multiple sclerosis (MS).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years, mean (SD); range</td>
<td>53.89 (12.27); 35-71</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Female</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Handedness, n (%)</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Left</td>
<td>0 (0)</td>
</tr>
<tr>
<td>MS type, n (%)</td>
<td></td>
</tr>
<tr>
<td>Relapsing-remitting</td>
<td>5 (56)</td>
</tr>
<tr>
<td>Primary progressive</td>
<td>2 (22)</td>
</tr>
<tr>
<td>Secondary progressive</td>
<td>2 (22)</td>
</tr>
<tr>
<td>EDSS(^a) score, mean (SD); range</td>
<td>3.89 (1.95); 2.0-7.5</td>
</tr>
<tr>
<td>Disease duration in years, mean (SD); range</td>
<td>10.56 (9.76); 1-28</td>
</tr>
<tr>
<td>AMSQ(^b) score, mean (SD); range</td>
<td>53.78 (21.38); 36-99</td>
</tr>
<tr>
<td>9HPT(^c) in seconds, mean (SD); range</td>
<td></td>
</tr>
<tr>
<td>Right + left</td>
<td>26.27 (8.71); 20.20-46.87</td>
</tr>
<tr>
<td>Right</td>
<td>24.17 (5.20); 18.67-35.50</td>
</tr>
<tr>
<td>Left</td>
<td>28.36 (14.60); 19.51-65.53</td>
</tr>
</tbody>
</table>

\(^a\)EDSS: Expanded Disability Status Scale. Scores range from 0 (normal neurological examination) to 10 (death due to MS) in 0.5-unit increments that represent higher levels of disability.

\(^b\)AMSQ: Arm Function in Multiple Sclerosis Questionnaire. Scores, for each of the 31 questions, range from 1 (not at all limited) to 6 (no longer able to), with total scores ranging from 31 to 186.

\(^c\)9HPT: Nine-Hole Peg Test.

Figure 2. Results of the System Usability Scale (SUS). The SUS score ranges from 0% to 100%, where a higher value indicates better system usability. The dotted line represents a score of 71.4%. A score of 70% up to a maximum 100% represents acceptable-to-excellent usability. PwMS: persons with MS; HS: healthy subjects.
Discussion

Principal Findings

We extensively evaluated the feasibility and usability of a tablet app–based dexterity home-based training program in both persons with MS and healthy subjects. The very high adherence rate toward the training protocol, being 97%, points to the excellent feasibility of the TAD-MS, which is further supported by the absence of dropouts. Regarding usability, both persons with MS as well as healthy subjects scored high on the SUS, were very satisfied with the system features of the app, and highly appreciated its easy integration into their daily life setting. Furthermore, our findings suggest that demographic factors (ie, gender; age; and disease duration, type, and severity) and dexterity levels do not interfere with the use of the app, further adding to the high usability of this mHealth intervention.

In contrast with the only two previous studies that tested the use of different dexterity apps in other populations, which were stroke survivors [18] and patients suffering from carpal tunnel syndrome [9], our pilot study shows for the first time a high adherence rate for a 4-week, tablet app–based dexterity training program in persons with MS. The duration and dosage of our training protocol, the TAD-MS, actually matched those from a previous dexterity app study [18] and is also in line with other home-based dexterity training programs in MS [32,33], which showed similar adherence rates. The fact that all included subjects did not need additional technical support at home, were independent, and were free to choose when to do their exercises (ie, morning, afternoon, or evening) likely contributed to the high adherence.

The user interface of this app was simple and clear, mainly explaining the high usability scores. Previous recent reviews [9,10] and surveys [7,8,34] about mHealth apps for MS already suggested some key features for the development of user-friendly apps. These features were as follows: easy app navigation, clear instructions of the exercises, aspects of gamification, and low technical requirements. This app fulfilled these key elements, since persons with MS could easily navigate through the app, with each exercise having clear instructions (ie, video sequences demonstrating and explaining the aim of the exercises). In addition, gamification elements and integrated feedback to optimize performance were incorporated, both aspects that are important to build and maintain motivation during training, as suggested before [9,34].

Regarding the effect of demographic variables on app usability scores, we did not find an association between the user interface and demographic variables, supporting the generalizability of the approach and indicating that the app is suitable for a broad MS population. This finding is in line with a previous report that found that irrespective of age, disease type, and disease severity, persons with MS were able to manage technical issues of a new mHealth app [35]. However, our results do contrast with Kizony and colleagues [17], who showed that declines in age-related dexterity, as previously shown in healthy subjects [36,37], could affect app performance. Thus, older healthy subjects were more challenged to use the app properly. The reason why we could not find such a relationship, besides the limited sample size, may lie in the heterogeneity of the MS disease, meaning that both younger and older patients, being differently affected by the disease, may have similar experiences using our app. The level of impaired dexterity, irrespective of age, might be a bigger challenge. In MS, a possible impact of impaired fine motor skills, mainly due to underlying ataxia, muscle weakness, and/or tremor [2], on app usability has been suggested as a possible barrier [7,34]. However, in our study, we could not find a relationship between validated dexterity measures and usability scores for this app, suggesting this might not be such a decisive barrier.

Limitations

This pilot study does have some limitations. We included a small sample of participants with relatively good cognitive function. Therefore, we cannot generalize our findings to individuals suffering from severe cognitive disabilities. However, this was not the scope of this pilot study. For persons with MS with severe cognitive impairment, more conventional dexterity-training methods might be more useful [32], because real objects and tools are used within this kind of training, certainly being more closely related to real dexterity ADL.
Finally, although we included a healthy control group, these subjects did not have any dexterous impairment and, therefore, did not participate in the 4-week training. Future studies could explore whether healthy older subjects with some dexterous disability could benefit from the TAD-MS.

Conclusions
This pilot study is the first to demonstrate the high feasibility and usability of a new tablet app–based dexterity home-based training program, called the TAD-MS [6], among both persons with MS and healthy subjects. The tablet app training program can be easily integrated into the home situation, without any technical or therapeutic support, therefore providing an additional tool in the rehabilitation toolbox to train dexterity for persons with MS. Whether the TAD-MS improves dexterity will need to be assessed in future studies (ie, in a pre-post design) that can provide effect sizes that are necessary to perform appropriate sample size calculations for randomized trials.

Acknowledgments
First, the authors would like to thank the Swiss Multiple Sclerosis Society and Bayer AG for their financial support. Second, we are also very grateful to all participants for their effort and valuable feedback. Third, we would like to thank Amy Orellana for providing her app. Additionally, we would like to show appreciation to Dr Dirk Lehnick for his methodological advice regarding statistics. Last, we are thankful to Cleo Bol for helping with data collection.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Custom User Engagement Questionnaire.

References


Abbreviations

- **9HPT**: Nine-Hole Peg Test
- **ADL**: activities of daily living
- **AMSQ**: Arm Function in Multiple Sclerosis Questionnaire
- **COGNI-TRAcK**: Cognitive Training Kit
- **CUEQ**: Custom User Engagement Questionnaire
- **EDSS**: Expanded Disability Status Score
- **EKNZ**: Northwest and Central Switzerland
- **mHealth**: mobile health
- **MS**: multiple sclerosis
- **PS**: prescribed number of sessions
- **Q**: question
- **SP**: number of home sessions performed
- **SUS**: System Usability Scale
- **TAD-MS**: tablet app–based dexterity home-based intervention in multiple sclerosis
- **TEAMS**: Tele-Exercise and Multiple Sclerosis
The Ubiquitous Cognitive Assessment Tool for Smartwatches: Design, Implementation, and Evaluation Study

Pegah Hafiz1,2, MSc; Jakob Eyvind Bardram1,2, MSc, PhD

1Digital Health Section, Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark
2Copenhagen Center for Health Technology, Kongens Lyngby, Denmark

Corresponding Author:
Pegah Hafiz, MSc
Digital Health Section
Department of Health Technology
Technical University of Denmark
Richard Petersens Plads
Building 324, 2nd Floor, Room 270
Kongens Lyngby, 2800
Denmark
Phone: 45 91858371
Email: pegh@dtu.dk

Abstract

Background: Cognitive functioning plays a significant role in individuals’ mental health, since fluctuations in memory, attention, and executive functions influence their daily task performance. Existing digital cognitive assessment tools cannot be administered in the wild and their test sets are not brief enough to capture frequent fluctuations throughout the day. The ubiquitous availability of mobile and wearable devices may allow their incorporation into a suitable platform for real-world cognitive assessment.

Objective: The aims of this study were threefold: (1) to evaluate a smartwatch-based tool for the assessment of cognitive performance, (2) to investigate the usability of this tool, and (3) to understand participants’ perceptions regarding the application of a smartwatch in cognitive assessment.

Methods: We built the Ubiquitous Cognitive Assessment Tool (UbiCAT) on a smartwatch-based platform. UbiCAT implements three cognitive tests—an Arrow test, a Letter test, and a Color test—adapted from the two-choice reaction-time, N-back, and Stroop tests, respectively. These tests were designed together with domain experts. We evaluated the UbiCAT test measures against standard computer-based tests with 21 healthy adults by applying statistical analyses significant at the 95% level. Usability testing for each UbiCAT app was performed using the Mobile App Rating Scale (MARS) questionnaire. The NASA-TLX (Task Load Index) questionnaire was used to measure cognitive workload during the N-back test. Participants rated perceived discomfort of wearing a smartwatch during the tests using a 7-point Likert scale. Upon finishing the experiment, an interview was conducted with each participant. The interviews were transcribed and semantic analysis was performed to group the findings.

Results: Pearson correlation analysis between the total correct responses obtained from the UbiCAT and the computer-based tests revealed a significant strong correlation ($r=0.78$, $P<0.001$). One-way analysis of variance (ANOVA) showed a significant effect of the N-back difficulty level on the participants’ performance measures. The study also demonstrated usability ratings above 4 out of 5 in terms of aesthetics, functionality, and information. Low discomfort (<3 out of 7) was reported by our participants after using the UbiCAT. Seven themes were extracted from the transcripts of the interviews conducted with our participants.

Conclusions: UbiCAT is a smartwatch-based tool that assesses three key cognitive domains. Usability ratings showed that participants were engaged with the UbiCAT tests and did not feel any discomfort. The majority of the participants were interested in using the UbiCAT, although some preferred computer-based tests, which might be due to the widespread use of personal computers. The UbiCAT can be administered in the wild with mentally ill patients to assess their attention, working memory, and executive function.

(JMIR Mhealth Uhealth 2020;8(6):e17506) doi:10.2196/17506

KEYWORDS

cognition; memory; response time; attention; Stroop task; wearable devices; mobile phone
**Introduction**

**Background**

Wearable devices provide an opportunity for users to collect their personal data. A recent empirical study determined that fashnology, individuals’ attitudes, and risk context were the most influential factors in adoption of wearable devices for quantified self-tracking purposes [1]. Wrist-worn devices, particularly smartwatches, are becoming more popular. Usefulness and visibility are the two major reasons that people adopt a smartwatch [2]. Smartwatches are lightweight and portable, which makes them easy for people to wear and use in almost every context, while some people may not carry their smartphones when they go for a walk or run. Moreover, platforms, including Fitbit OS (operating system), Apple Watch’s watchOS, and Google’s Wear OS, support building stand-alone apps that run without connecting to a smartphone. The application programming interface (API) of some smartwatches allow sensor data collection in the wild, including physiological and behavioral data, such as sleep, heart rate variability, mobility, and location. King and Saffarzadeh reviewed the application of smartwatches in 27 health-related studies [3]. Their findings show that activity monitoring, chronic disease self-management, nursing or home-based care, and health care education are the current smartwatch-based applications in health care. Hence, smartwatches are suitable devices to assist researchers in developing stand-alone health care-related apps, as well as for collecting sensor data in the wild.

Cognitive functioning is a crucial aspect of mental health and determines the quality of individuals’ daily activities. According to Lyon et al, impairment in attention, memory, and executive function may cause problems at school or work [4]. Moreover, previous studies have shown daily fluctuations in alertness [5], working memory [6], and executive skills [7]. Quantifying cognitive performance may help individuals reflect on their own fluctuations. For instance, students can track their alertness levels to select appropriate times of day to schedule their attention-demanding tasks. Besides healthy individuals, mentally ill patients also suffer from cognitive dysfunction, such as dementia [8], bipolar disorder [9,10], attention deficit hyperactivity disorder (ADHD) [11], and schizophrenia [12]. Monitoring cognitive performance can thus help patients in scheduling their follow-up visits in case of significant degradation in their cognitive functioning, as it may indicate the onset of their illness.

Digital cognitive screening tools have been designed for different technological platforms, targeting both mentally ill patients and healthy individuals. The Cambridge Neuropsychological Test Automated Battery (CANTAB) Mobile [13], the Internet-based Cognitive Assessment Tool (ICAT) [14], the THINC-integrated tool (THINC-it) [15], MyCognition Quotient (MyCQ) [16], CogState [17], and the Brief Assessment of Cognition in Schizophrenia (BACS) [18] are some examples of the validated cognitive test batteries administered on a computer or tablet. The existing cognitive test batteries are administered at a certain time in a controlled condition. Such cognitive tools are not feasible for long-term frequent monitoring and assessment of cognitive functioning, since (1) it takes at least 15 minutes to complete a set of tests and (2) the tests are taken in a controlled condition without any distraction, for example, a silent room. However, according to previous studies [19,20], it is crucial to assess cognitive functioning in real-life settings for frequent and continuous monitoring of the individuals.

Ecological momentary assessment (EMA) [21] and the experience-sampling method (ESM) [22] were developed to overcome the bias in delivering retrospective self-reports by study participants. Both methodologies provide an opportunity to collect psychological and clinical measures of behavior, cognition, and emotion in situ [23]. Unobtrusive cognitive tests instead of subjective ratings may improve the accuracy of EMA and the ESM in longitudinal studies.

Taking together, a ubiquitous tool providing continuous and frequent assessment of the individuals’ in-the-wild cognitive performance would be an important approach for real-world psychometric research and diagnosis.

**Previous Studies**

**Overview**

The application of neuropsychological tests on mobile platforms previously showed promising outcomes [19,24]. In this section, an overview of the previous studies on digital cognitive tests developed for smartphones and smartwatches is presented. Commercial cognitive training mobile apps with no evidence of validity were excluded.

**Smartphone-Based Tools**

A research platform called iVitality includes a smartphone app with five cognitive tests, namely Memory-Word, Trail Making, Stroop, Reaction Time, and N-back. Jongstra et al conducted a study with 151 healthy individuals to examine feasibility and validity of the iVitality platform over 6 months [25]. According to the results of their validation study, the Stroop and Trail Making tests correlated moderately ($r=.5$ and $r=.4$, respectively) with the conventional tests. The authors did not validate the rest of the cognitive tests against their corresponding baseline measures, due to the difference between the raw scores of the smartphone tests and conventional tests. The Color-Shape Test (CST) is a smartphone-based app designed to measure cognitive processing speed and attention in the elderly population. The validity of the CST was examined against the Uniform Data Set (UDS) neuropsychological test battery in an experiment by Brouillette et al with 57 individuals who did not have dementia [26]. Their findings showed a significant correlation between CST scores and global cognition with the Mini-Mental State Examination ($r=.52$), Digit Span ($r=.43$), the Trail Making test ($r=.65$), and the Digit Symbol test ($r=.51$). However, the CST scores did not correlate with verbal fluency tasks. Tieges et al conducted a study with 20 delirium patients to assess the feasibility of a smartphone-based app called the DelApp against a computerized device called the Edinburgh Delirium Test Box (EDTB) [27]. The authors found no significant difference between the scores of the DelApp and the EDTB ($P=.41$). Pal et al used a mobile app called the Neurophone, which includes...
N-back, Stop Signal, and Stroop tests, to evaluate the cognitive performance of 20 healthy and 16 methamphetamine users against a validated computerized tool [28]. The Stop Signal test results could not be compared to the computerized tests due to the different parameters used by phone- and computer-based tests, while the scores of the N-back test on both platforms were similar. The authors used speech recognition in the Stroop test of their mobile app to detect the correct response time (RT). However, due to the inaccuracy of the speech recognition, the test results of the computer- and phone-based tests were not comparable. Dingler et al developed a smartphone-based tool including three short cognitive tasks, namely the psychomotor vigilance task (PVT), the go/no-go task, and the multiple object tracking task [29]. The authors conducted an in-the-wild study to assess the alertness of 12 participants over 9 days, on average. Although the short version of the PVT was validated before by Basner et al [30], the go/no-go and the multiple object tracking tasks were not tested against a computer- or paper-based neuropsychological test.

Smartwatch-Based Tool
Cormack et al developed a tool called the Cognition Kit for the Apple Watch, including a variation of the N-back test adapted from CANTAB’s N-back along with self-reports of mood using a short questionnaire [31]. The authors conducted feasibility and validation studies of the Cognition Kit with 30 depressed patients. According to their validation study results, N-back test performance correlated with CANTAB’s rapid visual information processing task ($P \leq .01, r=.5$).

Gaps in the Literature
The study conducted by Dingler et al [29] was the only work that introduced a smartphone-based toolkit for doing research in situ alertness. The rest of the smartphone-based apps were developed to deliver personal cognitive assessment tools without collecting mobile data. So far, the Cognition Kit is the only smartwatch-based tool exclusively assessing working memory through the N-back test. A limited number of cognitive measures provided by mobile tools, as well as a lack of studies in exploring the potential of smartwatches in measuring in-the-wild cognition, led us to build the Ubiquitous Cognitive Assessment Tool (UbiCAT). Our tool has three smartwatch-based cognitive tests measuring three key cognitive domains, namely attention, working memory, and executive function. The UbiCAT tests, along with smartwatch-based sensor data collection, allow researchers to analyze associations between individuals’ cognitive, physiological, and behavioral features toward identifying digital biomarkers of human cognitive functioning and conducting psychometric research in the wild.

Goals of This Study
Through this study, we will (1) evaluate the cognitive measures of the UbiCAT apps against state-of-the-art computer-based tools, (2) assess the usability of the UbiCAT tests, and (3) understand participants’ perceptions about smartwatch apps for assessing cognition.

Methods
In this section, we first provide details of the design and functionality of the UbiCAT apps; we then explain the study in detail.

Design Methods
Overview
The UbiCAT includes three smartwatch-based apps; each is a cognitive test that measures a certain cognitive domain. We considered three inclusion criteria for the UbiCAT tests: (1) the tests should measure memory, attention, and executive function, since fluctuations in these domains may negatively affect individuals’ work or study performance, (2) each test should be able to be adapted for the limited screen size of the smartwatch, and (3) each test should not require a microphone or speaker, which are essential in verbal recall tests. Taking these together, we selected a two-choice reaction-time test [32] to measure attention, the Stroop color-word test [33] to measure attention and executive function, and the N-back test [34] to examine working memory. The three tests contribute to short assessments, as it takes approximately 5 minutes to take the UbiCAT tests.

Three experts who each hold a doctoral degree within cognitive psychology and human-computer interaction were involved in the design process. First, the initial design of the aforementioned tests was sketched on paper. Based on detailed analysis of the available smartwatch hardware platforms, the Fitbit Ionic device was selected. Second, functional prototypes for each test were implemented separately and tested on the smartwatch. Individuals with different finger sizes were asked to work with the apps to adjust the size of the app buttons and text. The Fitbit design guidelines were also considered during the prototyping phase. The components of the UbiCAT apps were revised several times after meetings with the domain experts. Overall, the design and implementation process took 4 months. Third, a formative evaluation study of the earlier versions of the UbiCAT apps was conducted with 5 participants aimed to examine the usability of the apps and understand participants’ adoption of wrist-worn devices [35]. The findings of the formative evaluation study helped us improve the user interface and functionality of the apps.

The UbiCAT Cognitive Tests
Overview
Three stand-alone apps were built for the Fitbit smartwatch. Each test takes less than 2 minutes to complete. We selected the following names for the UbiCAT apps to simplify memorizing the apps for the users: Arrow test (two-choice reaction-time test), Letter test (N-back test), and Color test (Stroop color-word test). An outline of the UbiCAT apps is presented in the following sections and snapshots are shown in Figures 1-3.

Arrow Test
The Arrow test presents a sequence of rightward or leftward arrows to the user one by one. The user is required to select the correct direction of each arrow by tapping on either the left or
right app button. The position of each arrow can be on the left or right side of the screen. Figure 1 shows a snapshot of the Arrow test where the correct response to this stimulus is the app button on the right side.

**Letter Test**

In the Letter test, a sequence of English alphabet letters are displayed to the user. Depending on the value of N, the user is supposed to determine whether the current stimulus is the same as the N letter, or N letters, back in the sequence or not. The value of N determines the difficulty level and is unchanged during an entire trial. Figure 2 shows a snapshot of the 2-back test, where N is equal to 2.

**Color Test**

The names of four colors, for example RED, with either the same or different ink color are the stimuli of the Color test. A congruent stimulus has the same color as its meaning, while an incongruent stimulus has a different color. The task of the user is to select the ink color of each stimulus by tapping on the app button labeled with the color name. Figure 3 presents an incongruent stimulus. Here, the correct response is the GREEN app button in the bottom-left corner.

---

**Figure 1.** A sample test taken from the UbiCAT (Ubiquitous Cognitive Assessment Tool) Arrow test. The stimuli is the rightward arrow and the app buttons on both sides capture the direction of the arrow.

**Figure 2.** A sample test taken from the UbiCAT (Ubiquitous Cognitive Assessment Tool) Letter test, 2-back task. The participant should indicate whether “T” appeared 2 letters back in the sequence or not.
**Technical Specifications and Apparatus**

Two validated computer-based tools, PsyToolkit [36,37] and the THINC-it application [15], were run on a MacBook Pro (15-inch Retina display, Apple Inc) during the study. A Fitbit Ionic smartwatch (1.42-inch screen, 348 × 250 pixel resolution) was used to run the UbiCAT apps. The figures in this paper were created in RStudio using the ggplot2 package [38].

**Ethical Approval**

The study protocol and system description were sent for approval by the Danish Ethical Committee. The study was classified as a nonclinical survey study and, hence, exempted for ethical approval (Journal-nr.: H-19086232).

**Participant Recruitment**

We recruited 21 healthy adults who lived in Copenhagen, Denmark, using a snowball sampling method [39]. All participants had sufficient English-language skills to read the test instructions. Participants were not eligible if they had a history of mental illness, were aged over 50 years, or had color blindness.

**Procedure**

**Overview**

All of the test sessions were performed in a silent room at the Technical University of Denmark. The study session lasted 60-75 minutes per participant. Participants were compensated with a gift card worth an amount equal to US $15 that was given at the end of the study. Prior to an experiment, the study leader (PH) informed the participant to ask for a short break between the testing sessions if needed. We measured each participant’s perceived wrist discomfort after completing each of the UbiCAT tests using a 7-point Likert scale. Figure 4 shows a participant completing a UbiCAT test on the Fitbit smartwatch. A detailed description of the experiment is presented below.

**Figure 3.** A sample test taken from the UbiCAT (Ubiquitous Cognitive Assessment Tool) Color test, displaying an incongruent stimuli.

**Figure 4.** A study participant completing a UbiCAT (Ubiquitous Cognitive Assessment Tool) test via a Fitbit Ionic smartwatch. The laptop was used to administer computer-based tests.
First, a general description of the study was given to the participant. Second, a consent form was handed to the participant. Upon signing the consent form, background information from the participant was collected, including age, gender, educational background, and preference in terms of watch-wearing wrist (ie, dominant or nondominant hand). Third, the participant was asked to perform the three UbiCAT tests one by one. Each test was administered against its corresponding computer-based test. PH explained the instructions of the computer-based tests to the participant and repeated if needed. The participant was able to read the instructions of the smartwatch-based tests in the UbiCAT by themself. The feedback displayed to the participant was the fraction of correct responses to the total responses in each UbiCAT test. The interaction of each participant with the smartwatch was video-recorded during the experiment. The order of test administration on the smartwatch and computer was counterbalanced among participants.

Previous work mentioned that some of the cognitive test results obtained from paper-based and computer-based tools could not be compared to their corresponding smartphone-based tests, due to the difference between parameters. Therefore, we chose the PsyToolkit for the Stroop color-word test and the N-back test. This tool allows researchers to program their experiments and adapt the parameters to their needs. We matched the difficulty levels of the N-back tests by changing the N and selecting the same ratio of congruent stimuli to the incongruent stimuli (1:3) in the Stroop tests. The details of our study are presented below.

**Arrow Test Versus Spotter Test**

All participants took the Arrow test and THINC-it Spotter test twice. The stimuli of each test on both the smartwatch and the computer was a set of 40 arrows. Each arrow was displayed on the watch for a maximum of 2000 ms. The interstimulus interval was randomly selected to be between 1000 and 3000 ms. The input of the THINC-it Spotter test was received by pressing the left or right arrow key, while the input was captured by tapping the left or right app button in the UbiCAT Arrow test. The performance measure calculated for both tests was the number of correct responses and fastest RTs.

**Letter Test Versus PsyToolkit N-Back Test**

The N-back test was administered separately with three difficulty levels, starting from N=1. The tests with the same difficulty level were tested against each other. For instance, 1-back in the Letter test was examined against the PsyToolkit 1-back test. The stimuli of each test was a sequence of 40 English alphabet letters displayed one by one. The time limit for the participant to respond to a stimulus was 2500 ms. Two keys were used to respond during the PsyToolkit test: “m” for yes and “n” for no. The inputs were captured on the UbiCAT Letter test by tapping on the app buttons labeled as “Yes” and “No.” The performance measures were the number of correct responses and mean RTs to the stimuli.

**Color Test Versus PsyToolkit Stroop Test**

All participants took each test twice. The stimulus of each test was 30 color names consisting of 7 congruent and 23 incongruent color names. The time limit was 2500 ms. Participants were required to press “b” for blue, “g” for green, “r” for red, and “y” for yellow in the PsyToolkit Stroop test. Responses were captured in the UbiCAT Color test by tapping on the app buttons labeled with the color names (see Figure 3). The pink color replaced yellow on the Fitbit smartwatch for some participants who found yellow difficult to distinguish. The performance measures of the Stroop tests were the mean RTs to the congruent and incongruent stimuli.

**Usability Testing**

The usability of the UbiCAT apps was assessed using the Mobile App Rating Scale (MARS) questionnaire [40]. Relevant questions concerning aesthetics, functionality, and information were selected from the MARS questionnaire (see Multimedia Appendix 1). The rating scale for each of the MARS questions ranged from 1 (the lowest score) to 5 (the highest score).

**Perceived Cognitive Workload**

Each N-back task was preceded by the NASA-TLX (Task Load Index) questionnaire [41] to quantify participants’ perceived cognitive workloads using a 7-point Likert scale. The following subscales of the NASA-TLX were used: mental demand, temporal demand, performance, effort, and frustration level. It should be noted that the physical demand subscale was excluded as it was deemed irrelevant.

**Follow-Up Interview**

Upon finishing each experiment, a short interview was performed with each participant to investigate their subjective perception about the experiment and the UbiCAT tests, as well as their suggestions to improve the apps and/or instructions. The interviews were audio-recorded and transcribed for semantic analysis and grouping of the findings across participants.

**Statistical Analysis**

The Pearson correlation test was performed on the number of correct responses and mean RTs of the cognitive tests on both platforms. The paired-sample t test was applied on the performance measures to compare the numbers obtained from the smartwatch- and computer-based tests. One-way analysis of variance (ANOVA) was used to analyze the effect of difficulty level on the participants’ test performances during the N-back test. The CI of the statistical tests was 95%. The statistical analysis was performed in JASP, version 0.11.1 (The JASP Team).

**Results**

**Participant Statistics**

Participants were aged between 19 and 44 years (mean 26, SD 6), and 9 out of 21 participants (43%) were female. On average, participants spent 5.7 years studying at a higher-education level. Participants had diverse occupational backgrounds, including design, computer science, water engineering, construction, health care, energy, and food engineering. Of the 21 participants, 10 (48%) of them had used at least one wrist-worn device before. All participants except for 1 (20/21, 95%) wore the smartwatch on their nondominant hand.
Overall Analysis

Pearson correlation analysis revealed a significant strong correlation between the total number of correct responses obtained from the cognitive tests on the UbiCAT and computer-based tools \((r=.78, P<.001)\). It should be noted that the scores of 4 participants of the PsyToolkit Stroop test were lost; thus, the correlation analysis between the total scores was performed for 17 participants. Figure 5 shows the total participant accuracy obtained from the UbiCAT apps versus the computer-based tools, along with the regression line. The single data point located on the bottom-left corner of Figure 5 might indicate an outlier; however, we did not remove this sample point, since it is normal that the abilities of the individuals are different from each other.

**Figure 5.** Overall participant accuracy in the three cognitive tests. Each black dot represents results from one participant. The blue line is the regression line and the shaded region is the CI. UbiCAT: Ubiquitous Cognitive Assessment Tool.

Two-Choice Reaction-Time Tests

The Pearson correlation analysis that was applied on the average of correct responses in the two trials of the Arrow test and Spotter test and the participants' fastest RTs on both platforms is presented in Table 1. Figure 6 shows the box plots of the number of correct responses for both platforms during each trial. Figure 7 shows the box plots of the participants' fastest RTs calculated for both trials of the two-choice reaction-time tests. We applied the paired-sample Student \(t\) test and it revealed that the fastest RTs obtained from the Arrow test in both trials were not statistically different \((t_{20}=-1.266, P=.22)\). The average of the participants' fastest RTs in the Arrow test were statistically higher than in the Spotter test \((t_{20}=10.84, P<.001)\).

N-Back Test

Figures 8 and 9 show the number of correct responses and the mean RTs of the participants, respectively, during the 1-back, 2-back, and 3-back tests in the Letter test and the PsyToolkit N-back test. Pearson correlation analysis was performed on the number of correct responses and the mean RTs for each difficulty level between the Letter test and PsyToolkit N-back test. The results are presented in Table 2.

**Table 1.** Correlation analysis between performance measures in the Arrow test and the Spotter test.

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Pearson (r)</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of correct responses</td>
<td>.61</td>
<td>.003</td>
</tr>
<tr>
<td>Fastest response times</td>
<td>.24</td>
<td>.30</td>
</tr>
</tbody>
</table>
Figure 6. Box plots of participants’ number of correct responses during the two-choice reaction-time tests.

Figure 7. Box plots of participants’ fastest response times during the two-choice reaction-time tests.
Figure 8. Box plots of participants’ number of correct responses in the N-back tests.

Figure 9. Box plots of participants’ mean response times during the N-back tests.
One-way ANOVA was performed to analyze the effect of difficulty level on the participants' test performances (see Multimedia Appendix 2).

The results of the NASA-TLX questionnaire for 1-back, 2-back, and 3-back on both platforms are reported in Table 3. The numbers in this table show the means and SDs calculated based on the 7-point Likert scales for the metrics of the NASA-TLX.

Table 2. Correlation analysis between performance measures of the N-back tasks in the Letter test and PsyToolkit N-back test.

<table>
<thead>
<tr>
<th>Performance measure and tasks</th>
<th>Pearson r</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean response time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-back</td>
<td>.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2-back</td>
<td>.71</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3-back</td>
<td>.53</td>
<td>.01</td>
</tr>
<tr>
<td>Number of correct responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-back</td>
<td>.90</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2-back</td>
<td>.19</td>
<td>.40</td>
</tr>
<tr>
<td>3-back</td>
<td>.35</td>
<td>.13</td>
</tr>
</tbody>
</table>

Table 3. The N-back cognitive workload results using the NASA-TLX (Task Load Index) metrics.

<table>
<thead>
<tr>
<th>Device and task</th>
<th>Score\textsuperscript{a} for each metric, mean (SD)</th>
<th>Frustration level</th>
<th>Effort</th>
<th>Overall performance</th>
<th>Temporal demand</th>
<th>Mental demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartwatch</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-back</td>
<td>2.81 (1.50)</td>
<td>2.05 (1.32)</td>
<td>2.52 (1.60)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-back</td>
<td>4.71 (1.27)</td>
<td>4.29 (1.49)</td>
<td>4.43 (1.17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-back</td>
<td>5.19 (1.33)</td>
<td>4.67 (1.62)</td>
<td>5.10 (1.10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-back</td>
<td>2.76 (0.99)</td>
<td>2.91 (1.84)</td>
<td>2.67 (1.07)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-back</td>
<td>4.50 (1.54)</td>
<td>3.10 (1.52)</td>
<td>4.35 (1.35)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-back</td>
<td>5.52 (1.29)</td>
<td>4.76 (1.76)</td>
<td>5.00 (1.18)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Scores were based on the 7-point Likert scales of the NASA-TLX metrics.

Stroop Color-Word Test

Figures 10 and 11 present the box plots of the mean RTs to the congruent and incongruent stimuli for each trial of the Color test and the PsyToolkit Stroop test, respectively. Table 4 reports the correlation analysis between the performance measures of the Stroop tests on both platforms. Box plots of the number of correct responses to both congruent and incongruent stimuli are shown in Figure 12.

Usability Ratings

The psychometric factors considered for the usability test were aesthetics, functionality, and information. Each of the UbiCAT apps were rated separately by the participants. Table 5 reports the means and SDs of the usability ratings, which are out of 5.
**Figure 10.** Box plots of participants' mean response times to congruent stimuli during Stroop tests.

**Figure 11.** Box plots of participants’ mean response times to incongruent stimuli during the Stroop tests.
Table 4. Correlation analysis between performance measures in the Color test and the Stroop color-word test.

<table>
<thead>
<tr>
<th>Performance measure</th>
<th>Pearson r</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean response times to congruent stimuli</td>
<td>.67</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Mean response times to incongruent stimuli</td>
<td>.66</td>
<td>.001</td>
</tr>
</tbody>
</table>

Figure 12. Box plots of participants’ number of correct responses in the Stroop tests.

Table 5. Usability ratings of the UbiCAT (Ubiquitous Cognitive Assessment Tool) apps.

<table>
<thead>
<tr>
<th>UbiCAT app</th>
<th>Scorea for each factor, mean (SD)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aesthetics</td>
<td>Functionality</td>
<td>Information</td>
</tr>
<tr>
<td>Arrow test</td>
<td>4.02 (0.76)</td>
<td>4.55 (0.52)</td>
<td>4.24 (0.86)</td>
</tr>
<tr>
<td>Letter test</td>
<td>4.19 (0.75)</td>
<td>4.36 (0.62)</td>
<td>4.33 (0.60)</td>
</tr>
<tr>
<td>Color test</td>
<td>4.14 (0.83)</td>
<td>4.64 (0.45)</td>
<td>4.31 (0.64)</td>
</tr>
</tbody>
</table>

aScores ranged from 1 (the lowest score) to 5 (the highest score).

Perceived Discomfort

For each UbiCAT app, participants rated the discomfort level in their wrist on which they wore the smartwatch via a 7-point Likert scale from 1 (the least discomfort) to 7 (the most discomfort). The corresponding means and SDs of the discomfort levels, calculated separately for the Arrow test, Letter test, and Color test, are 2.71 (SD 1.79), 2.24 (SD 1.18), and 2.14 (SD 1.32), respectively.

Interviews

Overview

Seven themes were extracted from the participants’ responses and a brief description of each theme is presented below. The participants’ quotes are presented in Multimedia Appendix 3.

Perceptions About the Experiment

Participants were asked to describe their feelings about the experiment. They were generally engaged in the experiment: 5 participants out of 21 (24%) mentioned that the experiment was “fun,” 3 (14%) said it was “good,” and 3 (14%) said it was “fine.” Only 1 participant out of 21 (5%) believed that the
Experiment was too long. The rest of the participants did not express their opinions or had to leave immediately after the experiment.

**Input Modality**

Participants compared the input modalities of the smartwatch and computer. Participants #1 and #3 (2/21, 10%) preferred the app buttons of the UbiCAT Color test to the keyboard in the Stroop test. Participants #9, #12, and #19 (3/21, 14%) felt more comfortable with the app buttons in general, and participant #14 (1/21, 5%) liked the tangibility of the keyboard.

**Device Screen**

Some participants compared the screen size of the smartwatch with the computer. Out of 21 participants, 1 (5%) argued that the bigger screen of the computer influenced his or her performance positively and 2 (10%) preferred the screen size of the computer to the smartwatch. We understood that computer screen size might be more acceptable for some people due to the longer adoption time of personal computers compared to smartwatches.

**Visual Impact**

Out of 21 participants, 3 (15%) implied that their better performance on the computer was due to the visualization of the elements. A participant (1/21, 5%) did not like the visual elements of the Fitbit, indicating that the overall device design and graphics affected participants' interaction quality apart from the specific user-interface design of the UbiCAT apps.

**Psychological Factors**

Apart from the physical characteristics of a smartwatch and a computer, psychological factors also influenced participants' performance. A participant (1/21, 5%) pointed to the gamified nature and playfulness of the UbiCAT tests.

**Performance**

Some of the participants related their lower performance in the UbiCAT tests to the apps. Out of 21 participants, 4 (19%) mentioned that the Color test sometimes did not capture their taps on the app buttons during the test. We noticed such an incident while reviewing the records of the experiments. The position of the app buttons in the Color test changed randomly to avoid practicing the positions of the buttons. It surprised some of the participants during the test. Besides, 1 of the participants (5%) thought that his or her performance might differ significantly between the first and second trials of the cognitive tests.

**Suggestions**

Out of the 21 participants, 3 of them (14%) proposed suggestions regarding the font size used in the UbiCAT tests.

**Discussion**

**Principal Findings**

UbiCAT implements three smartwatch-based cognitive assessment tests for in-the-wild deployment. The findings of this study revealed comparable performance measures to computer-based tests. The strong correlation between the overall accuracy of the participants during the cognitive tests in the UbiCAT and computerized tools showed that UbiCAT can be utilized for assessing individuals' three key cognitive functions, namely attention, working memory, and executive function. The analysis between the following performance measures of the UbiCAT and computerized tests revealed significant correlation coefficients: the number of correct responses in the two-choice reaction-time test; mean RTs in the 1-back, 2-back, and 3-back tests; the number of correct responses in the 1-back test; and the mean RTs to the Stroop test's congruent and incongruent stimuli.

The psychometric factors, including aesthetics, functionality, and information quality and quantity, of the UbiCAT apps had high average ratings by the participants (>4 out of 5). The subjective ratings of the participants' wrist discomfort levels were less than 3 out of 7, indicating that interaction with the UbiCAT apps via the smartwatch was comfortable, which is in line with our overall objective of making cognitive assessment as simple and convenient as possible.

Previous work reported mobile cognitive test results along with paper-based or computerized tests. Comparison between the correlation coefficients reported in previous studies and in our study is not possible due to different parameters, number of participants, and target population. Nevertheless, our test outcomes obtained from computer- and smartwatch-based apps were comparable to each other, unlike some of the previous studies (eg, Neurophone Stop Signal test) that could not compare their results with computerized or paper-based tests due to dissimilar parameters.

**Two-Choice Reaction-Time Test Outcomes**

The average number of correct responses obtained from the THINC-it and Arrow tests correlated significantly with each other. As it can be seen in Figure 6, the majority of the participants received the highest score on both platforms, which may indicate a ceiling effect. The participants' fastest RTs, however, did not correlate with each other, which might be due to the different interaction methods on both platforms. The app buttons in the Arrow test (see Figure 1) disappeared on receiving an input or time-out until the next stimulus appeared, since an accidental tap on the buttons could impede calculating the real performance of the participants. We observed that the participants moved their index fingers away after tapping on an app button in the Arrow test, while they kept their fingers on the arrow keys on the computer keyboard during the THINC-it Spotter test. Such a difference between the users' interactions may explain the longer RTs of the UbiCAT Arrow test as compared to the THINC-it Spotter test. Nevertheless, the difference between the fastest RTs of the participants helped us in understanding the impact of interaction methods.

The fastest RTs measured via both platforms may indicate that the thresholds of individuals' alertness vary on the computer and smartwatch platforms. In our study, the average fastest RTs of the participants in the Arrow test was 545 ms (SD 88), while the corresponding result for the THINC-it Spotter test was 315 ms (SD 59). A study on the development of a brief version of the PVT (PVT-B) showed that 500 ms might be the threshold for an impaired alertness [30], which is in line with the average
fastest RTs obtained from the THINC-it computer-based test. However, the participants of the PVT-B study pressed a button to respond during the tests, which is similar to the interaction method of the THINC-it Spotter test. Therefore, this threshold may not be comparable to the fastest RTs obtained from the Arrow test on the smartwatch. To infer the level of impairment on the basis of the user’s fastest RT delivered via a smartwatch, a larger study is required that would include both healthy controls and cognitively impaired patients. Nevertheless, the findings of our study revealed that the average fastest RT of the healthy subjects to a smartwatch-based test is above 500 ms.

According to Figure 7, participants’ fastest RTs were almost the same during the first and second trials of the Spotter test (327 ms and 303 ms, respectively), while their responses were a bit slower in the second trial of the Arrow test (564 ms) compared to the first trial (526 ms). A paired-sample t test showed that the fastest RTs received from the Arrow test in both trials were not statistically different.

N-Back Test Outcomes

One-way ANOVA showed the effect of difficulty level on the number of correct responses and mean RTs in the UbiCAT Letter test and the PsyToolkit N-back test. The perceived cognitive workload in the N-back tests also revealed that as N-back tasks became more difficult, the participants’ cognitive workload increased. The mean RTs obtained from each N-back task on both the smartwatch and the computer correlated significantly. Figure 9 shows that the mean RT during the UbiCAT 2-back test was higher than that of the 3-back test, while statistical analysis revealed no significant difference between the mean RTs of the UbiCAT 2-back and 3-back tests. The RTs of the PsyToolkit 2-back and 3-back tests were not statistically different either (P > .99). According to Table 3, higher temporal effort reported through NASA TLX questionnaires for the 2-back Letter test compared to the 3-back test may imply that participants were more rushed during the 2-back test. Moreover, participants might have spent more time on practicing the 2-back test right after taking the 1-back test to adapt their mental skills, since the reported mental effort for both 2-back and 3-back tests were higher than for the 1-back test on both the computer and the smartwatch.

According to Table 2, the correlation analysis between the number of correct responses of the N-back tests on the smartwatch and the computer was only significant for the 1-back test. The lack of a significant correlation between the 2-back and 3-back tasks might be due to the N-back test itself, since the letter sequences of the N-back test were generated randomly and the maximum number of matches (ie, hits) during the N-back tests was not controlled to be the same between the computer and the smartwatch.

Stroop Test Outcomes

The RTs to the congruent and incongruent stimuli on the PsyToolkit and Color tests significantly correlated with each other. According to Figures 10 and 11, the RTs obtained from the second trials were lower than the first trials for both the PsyToolkit Stroop test and Color test. However, the magnitude of difference between the RTs in both trials of the Color test was lower than that in the PsyToolkit Stroop test. It might be due to the difference between the interaction methods of the tests. In the Color test, the order of app buttons was shuffled after a test run to avoid practicing the positions and increasing engagement with the apps. On the other hand, the position of the keys was obviously stable during the PsyToolkit tests. Hence, participants might get used to the position of the keys and respond faster in the second trial of the PsyToolkit Stroop test, while the changing position of the app buttons in the Color test took some time for them to practice with the new positions. The change in the position of the app buttons was intended to obtain reliable outcomes during future studies for longitudinal frequent administration.

Figure 12 shows that several participants received the highest score in the PsyToolkit Stroop tests (ie, 10 participants in trial 1 and 11 participants in trial 2), while the scores are more distributed in the Color tests (ie, 2 participants received the highest score in trial 1 and 6 participants in trial 2). In addition, we observed that sometimes the app buttons in the Color test did not capture touch inputs by the participants and some participants reported this issue during the interviews. Therefore, lower scores in the Color test might be due to the Fitbit’s touch sensitivity.

Perceptions From the Interviews

Seven themes were identified from the follow-up interviews with the participants. Some of the participants generally felt more comfortable when taking a cognitive test on the smartwatch compared to the computer, while some did not. Factors related to the physical aspects of the device, including the screen size and distance and the input modalities, affected their interactions. We understood that longer adoption times of computers compared to smartwatches may explain why some participants preferred computer tests to the UbiCAT apps. Therefore, deploying smartwatches into individuals’ daily lives may take some time and may not be useful for all. Psychological factors were also involved in determining participants’ engagement with UbiCAT, such as the gamified features of the tests.

Implications for Future Work

On the basis of our interviews, we decided to (1) add customized badges to the UbiCAT apps depending on participants’ test performances to motivate them toward continuous usage of the UbiCAT, (2) increase the font size of the stimulus in the Letter test since it was not easy for some of the participants to read, and (3) keep the right and left app buttons of the Arrow test on the screen after they tap on a button.

This study was conducted to evaluate our novel smartwatch apps against their corresponding computer tests, as well as to investigate participants’ perceptions about the study and usability of the UbiCAT apps. One of the future directions of the UbiCAT project is to identify digital biomarkers of human cognition. In an upcoming study, we will collect participants’ mobile data, including physiological and behavioral data, along with assessing their daily cognitive functioning through the UbiCAT apps to determine digital biomarkers of human cognition.
digital biomarkers would help researchers in building predictive models of individuals' cognitive impairment using mobile data. Sleep-stage data and heart rate variability (HRV) are the physiological data that can be collected through the Fitbit API. Sleep disturbance is, for instance, prevalent in bipolar patients [42]. The negative impact of poor sleep on mood and cognitive functioning is particularly noticeable in bipolar patients [43]. Fitbit smartwatches collect sleep duration and stages, which can help us in measuring the impact of sleep quality on next-day cognitive performance. Literature suggests a relationship between reduced HRV and impairment in inhibition control [44]. Moreover, reduced HRV was observed in bipolar and schizophrenic patients compared with healthy controls [45].

Our outlook is to create a Cognitive Watch to extend human knowledge regarding cognition.

Limitations
The limitations of this study are threefold. First, this study was conducted with 21 healthy adults who were recruited mostly from the campus of the Technical University of Denmark. This was deemed appropriate for evaluating the UbiCAT as compared to existing tools. However, studies involving patients and people with cognitive impairment are needed and are the focus of our upcoming studies. Second, the UbiCAT is designed for in-the-wild administration and, yet, this study was conducted in an indoor environment. This was done because cognitive performance fluctuates and in order to be able to assess cognition using both UbiCAT and the computer-based tests, the tests had to be administered right after each other on both platforms in order to achieve comparable measures. Therefore, moving the participants inside and outside between computer-based and smartwatch-based test sessions could yield unreliable cognitive measures for our study. In our upcoming studies, however, the UbiCAT will be used outside the clinic in order to collect real-world cognitive performance, which will be compared with cognitive assessments performed in a clinic. Third, the results indicated that the UbiCAT tests may not reflect the optimal performance of the participants compared to the computer-based tests. Nevertheless, frequent tests with the UbiCAT in upcoming studies and with various patient groups may better verify the optimal performance of the UbiCAT users.

Conclusions
In this study, the UbiCAT as a smartwatch-based tool for cognitive assessment was evaluated against computer-based cognitive assessment tools. The results revealed significant correlations between the total scores of the UbiCAT tests and standard computer-based tests. The psychometric factors regarding the aesthetics, functionality, and information quality and quantity of the apps yielded high usability ratings from the study participants. The majority of our study participants felt comfortable when using the UbiCAT. The findings of this study showed that the UbiCAT can be used for assessing attention, working memory, and executive function across participants' everyday lives, along with mobile data collection. Future studies can administer the UbiCAT to mentally ill patients to collect their daily cognitive functioning data and to compare their results with lab-based studies.

Acknowledgments
This research is funded by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement (No. 722561). We wish to thank our study participants.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Usability evaluation questionnaire.
[PDF File (Adobe PDF File), 68 KB - mhealth_v8i6e17506_app1.pdf ]

Multimedia Appendix 2
Analysis of variance on the N-back test results.
[PDF File (Adobe PDF File), 30 KB - mhealth_v8i6e17506_app2.pdf ]

Multimedia Appendix 3
Themes and their corresponding quotes extracted from the interviews.
[PDF File (Adobe PDF File), 47 KB - mhealth_v8i6e17506_app3.pdf ]

References


Abbreviations

ADHD: attention deficit hyperactivity disorder
ANOVA: analysis of variance
API: application programming interface
BACS: Brief Assessment of Cognition in Schizophrenia
CANTAB: Cambridge Neuropsychological Test Automated Battery
CST: Color-Shape Test
EDTB: Edinburgh Delirium Test Box
EMA: ecological momentary assessment
ESM: experience-sampling method
HRV: heart rate variability
ICAT: Internet-based Cognitive Assessment Tool
MARS: Mobile App Rating Scale
MyCQ: MyCognition Quotient
NASA TLX: NASA Task Load Index
OS: operating system
PVT: psychomotor vigilance task
PVT-B: brief version of the psychomotor vigilance task
RT: response time
THINC-it: THINC-integrated tool
UbiCAT: Ubiquitous Cognitive Assessment Tool
UDS: Uniform Data Set

©Pegah Hafiz, Jakob Eyvind Bardram. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 01.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
The Mediating Role of Organizational Reputation and Trust in the Intention to Use Wearable Health Devices: Cross-Country Study

Funmi Adebesin¹, BSc, MSc, PhD; Revingstone Mwalugha¹, BCom, BCom (Hons)
Department of Informatics, University of Pretoria, Pretoria, South Africa

Corresponding Author:
Funmi Adebesin, BSc, MSc, PhD
Department of Informatics
University of Pretoria
Corner of Lynwood and Roper Street
Hatfield
Pretoria, 0083
South Africa
Phone: 27 0124205667
Email: funmi.adebesin@up.ac.za

Abstract

Background: The use of consumer wearable health devices for fitness tracking has seen an upward trend across the globe. Previous studies have shown that trust is an important factor in the adoption and use of new technologies. However, little is known about the influence of organizational reputation and trust on the intention to use wearable health devices.

Objective: This study aimed to investigate the mediating role of organizational reputation and trust in the intention to use wearable health devices and to examine the extent to which the country of residence influenced the effect of organizational reputation on consumers’ trust in and intention to use wearable health devices.

Methods: We conducted a cross-country survey with participants from Kenya and South Africa using a Google Forms questionnaire derived from previously validated items. A series of mediation regression analyses were carried out using the PROCESS macro with the bootstrap CI procedure. A one-way, between-group multivariate analysis of variance (MANOVA) was also used to determine the key factors that distinguish Kenyans and South Africans in their intention to use wearable health devices.

Results: A total of 232 questionnaire responses were collected. The results revealed that organizational reputation significantly mediates the relationship between trust propensity and trust, with an indirect effect of 0.22 (95% CI 0.143-0.309). Organizational reputation also plays a significant direct role in the intention to use a wearable health device, with a direct effect of 0.32 (95% CI 0.175-0.483). This role is regardless of participants’ country of residence. Furthermore, there is a significant mediating effect of trust on the relationship between trust propensity and the intention to use a wearable health device, with an indirect effect of 0.26 (95% CI 0.172-0.349); between perceived security and the intention to use a wearable health device, with an indirect effect of 0.36 (95% CI 0.255-0.461); and between perceived privacy and the intention to use a wearable health device, with an indirect effect of 0.42 (95% CI 0.282-0.557). The MANOVA test shows statistically significant differences in all variables for both groups, with the exception of organizational reputation where there is no significant difference between the two cohorts.

Conclusions: Organizational reputation has a significant direct influence on participants’ trust in and the intention to use a wearable health device irrespective of their country of residence. Even in the presence of perceived security and perceived privacy, trust has a significant mediating effect on the intention to use a wearable health device.

(JMIR Mhealth Uhealth 2020;8(6):e16721) doi:10.2196/16721

KEYWORDS

fitness trackers; intention; Kenya; physical activity; privacy; South Africa; trust; regression analysis
Introduction

Background

Wearable health technologies continue to rank among the top 3 global fitness trends since 2016, maintaining the first position in 2016, 2017, and 2019 [1]. There is growing evidence that increasing physical activity (PA) is beneficial to personal well-being [2-5]. Several studies reported on the positive relationship between health consciousness and increase in PA [6-8]. Wearable health devices that monitor PA can take the form of an accelerometer, a pedometer, a heart rate monitor, or a combination of these mechanisms [9-12]. As a result of their capability to seamlessly monitor and capture health data [13], wearable health devices have become popular and valuable tools for monitoring and recording PA and other health-related data [10,14,15].

Consumer wearable health devices (CWHDs), such as Apple Watch, Fitbit, Garmin, and Huawei, typically come with a range of features that support real-time tracking of PA. These devices combine various behavioral change techniques, including goal setting, self-monitoring, feedback, and reward [10], to promote positive health habits. In many instances, a wearable health device is linked to one or more compatible mobile phone apps. CWHDs have the potential to help users improve their PA and adopt healthy behavior [10,16]. Through features such as goal setting, performance monitoring, and personalized feedback, users can set up fitness goals that meet their needs, track activities against set goals, and adjust targets as required [16,17]. The virtual fitness trainer feature in CWHDs and mobile fitness apps can mimic a personal fitness coach or trainer by showing users the correct body movements for a specific exercise [18,19]. This feature is particularly beneficial in the Global South context (countries classified as low- and medium-income countries in Africa, Asia, Latin America, and the Caribbean), where many citizens do not have the financial means to use the services of a professional fitness trainer.

Despite the potential benefits of CWHDs in supporting increased PA, there are legitimate concerns over privacy and the security of data collected by these devices [20,21]. Data collected by CWHDs are often transmitted to cloud storage. These data are at risk of unauthorized access by people with malicious intent on the device itself, while the data are in transit (and or) on cloud storage [22].

Several studies have been published on the use of CWHDs to track PA [23-26]. However, as acknowledged by Wiesner et al [21], few studies have focused on the privacy issues that influence the adoption of these devices. Even rarer are studies by researchers from Global South countries focusing on the privacy and security of CWHDs. To address this gap, we investigated the mediating role of 2 variables (ie, organizational reputation and trust) in the following relationships:

- The mediating role of organizational reputation in the relationship between trust propensity and trust.
- The mediating effect of trust in the relationship between trust propensity, perceived security, and perceived privacy in the intention to use wearable health devices.
- The mediating role of trust in the relationship between organizational reputation and the intention to use a wearable health device.

The following research questions are addressed in the paper:

- To what extent is the relationship between organizational reputation and the intention to use a wearable health device mediated by trust?
- To what extent are the relationships between the 3 factors (ie, trust propensity, perceived security, and perceived privacy) and the intention to use a wearable health device mediated by trust?
- What are the key factors that distinguish Kenyans and South Africans in their intention to use wearable health devices?

Previous studies, such as Gao et al [27], Gu et al [28], and Meyer et al [29], on the factors that influence the adoption and use of wearable health devices did not consider the mediating role of factors such as organizational reputation. Furthermore, studies such as Sillence et al [30] that specifically investigated the influence of organizational reputation on women’s trust in a website were not from the perspective of mediation. Another unique feature of our study is the fact that it is a cross-country study, which allowed us to collect data from 2 developing countries that are culturally diverse, yet with stark similarities [31,32]. Hence, this paper contributes to the field of ubiquitous health (uHealth) as it presents empirical research findings on the effect of organizational reputation on consumers’ trust in wearable health devices and the intention to use wearable health devices. In the following subsections, we discuss the theoretical background upon which the constructs that were tested in the study are based.

Trust Propensity and Trust

Trust propensity relates to an individual’s natural predisposition to trust another person. This trust is without regard for the circumstances, even with little or no previous information about the other person (or thing) [33,34]. A person (trustor) is said to be trusting of another party (trustee) when the trustor is prepared to be subjected to the consequences of the trustee’s actions. This willingness is based on the expectation that the trustee will behave as expected by the trustor regardless of whether the trustor has power to control the actions of the trustee [33]. Trust consists of 4 components, namely (1) competence, which is the belief that the other party is capable of discharging their responsibility to the trustor; (2) benevolence, the belief that the trustee is willing to act in the best interest of the trustor; (3) integrity, the belief that the trustee will be truthful, abide by agreements, act ethically, and fulfill promises made; and (4) predictability, the belief that over time, the actions of the other party will be consistent and the trustee can predict the trustee’s action in any given situation [35]. Trust propensity can influence the development of trust in a person or thing. Cultural background, developmental experiences, and personality types are thought to influence the willingness of an individual to trust other people or things [33].

Earlier studies have shown that trust propensity can shape consumers’ trust in the use of new technologies. For example, Zhou [36] found that trust propensity has a significant effect on
initial trust in and the intention to use mobile banking apps. In her study on the role of trust in citizens’ adoption and use of electronic government (eGovernment), Colesca [37] reported a positive relationship between trust propensity and trust in eGovernment, whereas Lee and Turban [38] also reported that trust propensity has a positive influence on consumers’ trust in electronic commerce (eCommerce) websites. Similarly, Gu et al [28] reported a positive link between trust propensity and initial trust in CWHDs.

Although trust is an important factor in the adoption and use of technologies such as mobile banking and eCommerce websites, it is even more the case when it comes to CWHDs that collect and analyze health data on a continuous basis [14]. Design issues such as accuracy, privacy, and security of health data can influence the level of trust and eventual intention to use technologies such as CWHDs [29,39].

Although previous studies, such as those by Gu et al [28], Zhou [36], Colesca [37], Lee and Turban [38], and Ribadu and Rahman [40], have shown the link between trust propensity and trust, little is known about the mediating role of organizational reputation in this relationship. To address this gap, we hypothesized the following:

- **H1:** The relationship between trust propensity and trust is mediated by organizational reputation.
- **H2:** The relationship between trust propensity and the intention to use a wearable health device is mediated by trust.

**Organizational Reputation and Trust**

Organizational reputation can be viewed from 2 perspectives, namely institutional and economic. From an institutional perspective, organizational reputation relates to the extent to which an organization creates value for its stakeholders and differentiates itself from its competitors [41,42]. From an economic perspective, organizational reputation is the ability of a company to produce quality goods and services [42]. Positive organizational reputation is generally seen as an intangible asset that can enable a company to grow its market share, maximize its profit, and attract new customers, while retaining existing ones [43-45].

Positive organizational reputation reflects the quality and performance of the products and (or) services being offered by an organization [46]. There is a direct link between organizational reputation and consumers’ trust and continued loyalty [41,47-52]. Although good organizational reputation is vital for building consumers’ trust in general, it is even more so in web-based services where customers are unable to feel or try out a product before making payments. Thus, customers may only have to rely on their gut feeling or the company’s reputation to inform the decision to interact with the web-based service [47,48].

Studies by Haery et al [47] and Jung and Seock [50] showed that positive organizational reputation influenced customers’ intention and continued loyalty. Flavían et al [52] also reported that organizational reputation has a significant effect on customers’ trust in web-based banking, compared with traditional banking methods. In a study that investigated the factors that influence women’s trust and mistrust in health websites, Sillence et al [30] also found that the majority of study participants were more trusting of information on health websites that is managed by reputable organizations. Thus, we can extend this relationship to consumers’ intention to use wearable health devices through the following hypothesis:

- **H3:** The relationship between organizational reputation and the intention to use a wearable health device is mediated by trust.

**Perceived Security and Trust**

Perceived security can be defined as the belief that one’s personal information will not be accessed, viewed, manipulated, or stored by unauthorized persons while in transit or on a storage device (location) [53]. The mechanisms that can be used to ensure the security of customers’ information include digital signatures, encryption, authentication, and verification [54-57].

The impact of perceived security on trust in new technologies has been studied by several researchers. For example, in their study on the adoption of internet banking, Patel and Patel [58] found that perceived security has a strong influence on participants’ intention to use internet banking services. Similarly, Aboobucker and Bao [59] found that security and privacy play a significant role in the adoption of internet banking. In another study on the acceptance of near-field communication (NFC)-based mobile payment by restaurant users, Khalilzadeh et al [60] found that perceived security influences customers’ trust in NFC-based mobile payment systems. In another study, Sharma et al [61] found that perceived security and perceived privacy have a significant influence on the trust in and subsequent use of social networks.

Authors of previous studies have reported on privacy and security concerns related to uHealth data [62-64], which is arguably more personal than biographical data such as name, date of birth, or telephone number. However, little is known about the mediating role of trust in the relationship between perceived security and the intention to use a wearable health device. Data collected by wearable health devices are at risk of unauthorized access by people with malicious intent on the device itself, while the data are in transit and (or) on cloud storage [22,65,66]. To test the mediating role of trust in the relationship between perceived security and the intention to use a wearable health device, we formulated the following hypothesis:

- **H4:** The relationship between perceived security and the intention to use a wearable health device is mediated by trust.

**Perceived Privacy and Trust**

Privacy can be defined as the right to determine or control what personal information is disclosed to other entities or third parties [67]. Concerns over the privacy of personal information are growing due to the increasing capability of new technologies to collect, process, and distribute large amounts of data [53]. Security measures provide technical assurance that, to keep personal data safe, best practices will be adhered to. Privacy, on the other hand, involves compliance with legal requirements.
and policies on how consumers’ personal data will be collected and used [53]. When asked to provide personal information on electronic platforms, a user will typically do a quick risk-benefit analysis. When the benefits are perceived to outweigh the associated privacy risks, the user is more likely to provide the requested information [27].

In their study on the acceptance of wearable technology in health care, Gao et al [27] found that one of the factors that influence the intention to adopt wearable health device for fitness tracking is perceived privacy risk. A high level of perceived privacy risk can have a negative influence on the trust in and intention to use new technologies. Conversely, a low level of perceived privacy risk will have a positive influence on consumers’ trust in and intention to use new technologies [28,61]. To determine the mediating role of trust in the relationship between perceived privacy and the intention to use a wearable health device, we proposed the following hypothesis:

- H5: The relationship between perceived privacy and the intention to use a wearable health device is mediated by trust.

We collected data from participants residing in two different countries, Kenya and South Africa. We cannot assume that participants from these countries are homogeneous, and the factors discussed in the preceding subsections would be applicable to the participants in the same way. Even for participants in the same country, this kind of an assumption would not be plausible. To investigate the key factors that distinguish participants from the two countries in their intention to use wearable health devices, we hypothesized the following:

- H6: There is a significant difference between Kenyans and South Africans in their intention to use wearable health devices.

Methods

Research Model and Measurements

The conceptual model of the relationships between trust propensity, organizational reputation, trust, perceived security, perceived privacy, and intention to use wearable health devices is illustrated in Multimedia Appendix 1. The questionnaire used in the study was based on constructs that have been validated and tested for their reliability by researchers as reflected in Multimedia Appendix 2. Items on the questionnaire were adapted to suit the purpose of our study through minor changes to their wordings. The items were measured using a 5-point Likert scale (1=strongly disagree to 5=strongly agree). The questionnaire went through a pretest process where 2 experts (one with experience in the field of electronic health and the other an expert in the design of quantitative studies) reviewed and made suggestions for its improvement. Following the modifications, a pilot study was conducted with 5 participants to test the usability of the questionnaire and determine if further modification is required. None of the pilot participants experienced problems understanding the statements on the questionnaire, thus no further changes were made before distributing the questionnaire.

To ensure that the minor changes made to items on the questionnaire did not negatively affect their reliability, we performed a Cronbach alpha coefficient test using SPSS version 25 (IBM Corp) to assess internal consistency among the items. The Cronbach alpha values of the constructs were as follows: trust propensity (.85), perceived security (.79), perceived privacy (.87), organizational reputation (.79), trust (.84), and intention to use wearable health devices (.77). Thus, changes to the wordings of items on our questionnaire did not have any negative impact on their internal consistency as the items were based on previously validated scales.

Study Design

A cross-country electronic survey was conducted with participants from Kenya and South Africa. The questionnaire was implemented using Google Forms and distributed over a period of 4 weeks between September and October in 2018. To reduce the possibility of multiple responses by the same person, the Google Forms were set up to accept only one response per device. As potential participants were not offered any reward, there was little incentive for a participant to submit more than one response using multiple devices. The link to the Google Forms questionnaire was sent to a diverse group of undergraduate students pursuing Bachelor of Commerce, Engineering, Education, and Law degrees at 3 universities in Kenya (University of Nairobi, Kenyatta University, and Moi University) and 1 in South Africa (University of Pretoria) via email and WhatsApp, a social media platform. As acknowledged by Topolovec-Vranic and Natarajan [68], social media platforms such as Facebook, WhatsApp, LinkedIn, and Instagram provide new opportunities to recruit study participants. For example, the recruitment of potential study participants on social media platforms offers many benefits, including global reach, snowballing effect, and fast distribution [69]. To increase the participation rate, we requested participants to forward the link to the questionnaire to other potential participants. Responses to the questionnaire were automatically captured on the Google Forms.

Ethical approval for the study was granted by the Research Ethics Committee at the School of Information Technology, University of Pretoria. Participation in the study was completely voluntary; no identifying data were collected; and participants’ responses were anonymous. All respondents gave consent to participate in the study.

Statistical Analysis Method

Data were analyzed using a mediation regression analysis and the between-group multivariate analysis of variance (MANOVA) test.

The mediation hypothesis analysis can be performed using the Monte Carlo CI, the Bayesian credible interval, the bootstrapping CI, or the Sobel methods [70]. However, the Monte Carlo CI, Bayesian credible interval, and bootstrapping CI tests have been shown to outperform the Sobel test [71-73].

To test hypotheses 1 to 5, we conducted a series of mediation regression analyses through bootstrapped CI using the PROCESS macro by Hayes [74]. PROCESS is a free, easy-to-use add-on for SPSS and statistical analysis system.
software. We used the PROCESS macro for SPSS to analyze our data. PROCESS comes with more than 70 predefined models [74]. We analyzed each of the 5 mediation hypotheses using the default model 4 in PROCESS. One of the benefits of the bootstrapped CI method for testing mediation is that it does not impose the assumption of a normal sample distribution. The method is also more robust when the sample size precludes the use of methods such as structural equation modeling [72].

H1 to H5 were tested using 1000 bootstrapped samples at 95% CI. To test H1 (the relationship between trust propensity and trust is mediated by organizational reputation), we ran the PROCESS macro with trust propensity as the independent variable, trust as the dependent variable, and organizational reputation as the mediating variable. Similarly, H2 (the relationship between trust propensity and the intention to use a wearable health device is mediated by trust) was tested by loading trust propensity as the independent variable, intention to use a wearable health device as the dependent variable, and trust as the mediating variable.

H3 (the relationship between organizational reputation and the intention to use a wearable health device is mediated by trust) was tested by loading organizational reputation as the independent variable, intention to use a wearable health device as the dependent variable, and trust as the mediating variable. We also tested H4 (the relationship between perceived security and the intention to use a wearable health device is mediated by trust) by loading perceived security as the independent variable, intention to use a wearable health device as the dependent variable, and trust as the mediating variable. Finally, H5 (the relationship between perceived privacy and the intention to use a wearable health device is mediated by trust) was tested by loading perceived privacy as the independent variable, intention to use a wearable health device as the dependent variable, and trust as the mediating variable.

The mediation regression analyses were carried out in 2 stages: (1) by loading all data from the 2 countries and (2) by separating the data according to participants’ country of residence. Separating the data enabled us to determine the extent of the differences in the mediation variables between the 2 cohorts.

To test H6 (there is a significant difference between Kenyans and South Africans in their intention to use wearable health devices), a one-way, between-group MANOVA test using SPSS was carried out to determine the factors that distinguish Kenyan and South African participants. The variables trust propensity, organizational reputation, trust, perceived security, perceived privacy, and intention to use a wearable health device were loaded as dependent variables, whereas country of residence was loaded as the independent variable.

To ensure that there is no serious violation of the underlying assumptions of MANOVA test, we tested for normality, linearity, univariate and multivariate outliers, homogeneity of variance-covariance, and multicollinearity. None of the underlying assumptions were violated, with the exception of the Levene test of equality of variances, where the variable, intention to use a wearable health device, has a significant value of $P=0.03$. This value is slightly lower than the recommended value of $P=0.05$.

We then used the Pillai trace to test for any statistically significant differences between the 2 groups. This is to ensure that the slight violation of equality of variance assumption did not influence the outcome of the MANOVA test. Pillai trace is known to be more robust than Wilks lambda when comparing groups in situations where there is some violation of the underlying assumptions of MANOVA [75].

Results

Demographics of Study Participants

A total of 232 responses were received. As shown in Table 1, there were 137 participants from Kenya and 95 from South Africa. A total of 58.2% (135/232) of participants were males, whereas 41.8% (97/232) were females. Within each country, 67.2% (92/137) of Kenyans were males, whereas 32.8% (45/137) were females. Gender distribution in South Africa was 45% (43/95) males and 55% (52/95) females.

Correlation Between Research Constructs

The results of the correlation between the constructs are presented in Table 2. The constructs have weak-to-strong positive relationships with each other at $P<.001$. There is a weak positive relationship between trust propensity and the other 5 constructs (ie, perceived security, perceived privacy, organizational reputation, trust, and intention to use a wearable health device). The relationship between perceived security and the constructs perceived privacy, organizational reputation, trust, and intention to use a wearable health device is medium to strong, with the strongest relationship being between perceived security and perceived privacy ($r=0.72$; $P<.001$). Perceived privacy also has a moderate-to-strong positive relationship with the constructs perceived security, organizational reputation, trust, and intention to use a wearable health device, with the strongest being between perceived privacy and trust ($r=0.72$; $P<.001$). Between trust and the constructs trust propensity, perceived security, perceived privacy, organizational reputation, and intention to use a wearable health device, the strongest relationship is with organizational reputation ($r=0.74$; $P<.001$). The correlation between intention to use a wearable health device and the other 5 constructs is weak to moderate, with the strongest occurring between intention to use a wearable health device and trust ($r=0.67$; $P<.001$).
Table 1. Descriptive statistics of study participants.

<table>
<thead>
<tr>
<th>Category and item</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (% per total sample)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>135 (58.2)</td>
</tr>
<tr>
<td>Female</td>
<td>97 (41.8)</td>
</tr>
<tr>
<td><strong>Gender per country, n (% within country)</strong></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>92 (67.2)</td>
</tr>
<tr>
<td>Female</td>
<td>45 (32.8)</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>43 (45)</td>
</tr>
<tr>
<td>Female</td>
<td>52 (55)</td>
</tr>
<tr>
<td><strong>Country of residence, n (% per total sample)</strong></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>137 (59.1)</td>
</tr>
<tr>
<td>South Africa</td>
<td>95 (40.9)</td>
</tr>
</tbody>
</table>

Table 2. Pearson correlation (r) matrix for trust propensity, perceived security, perceived privacy, organizational reputation, trust, and intention to use a wearable health device at \( P < .001 \) (n=232).

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Trust propensity</th>
<th>Perceived security</th>
<th>Perceived privacy</th>
<th>Organizational reputation</th>
<th>Trust</th>
<th>Intention to use a wearable health device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust propensity</td>
<td>N/A(^a)</td>
<td>0.398</td>
<td>0.359</td>
<td>0.325</td>
<td>0.382</td>
<td>0.258</td>
</tr>
<tr>
<td>Perceived security</td>
<td>0.398</td>
<td>N/A</td>
<td>0.720</td>
<td>0.494</td>
<td>0.598</td>
<td>0.478</td>
</tr>
<tr>
<td>Perceived privacy</td>
<td>0.359</td>
<td>0.720</td>
<td>N/A</td>
<td>0.620</td>
<td>0.721</td>
<td>0.542</td>
</tr>
<tr>
<td>Organizational reputation</td>
<td>0.325</td>
<td>0.494</td>
<td>0.620</td>
<td>N/A</td>
<td>0.742</td>
<td>0.620</td>
</tr>
<tr>
<td>Trust</td>
<td>0.382</td>
<td>0.598</td>
<td>0.721</td>
<td>0.742</td>
<td>N/A</td>
<td>0.671</td>
</tr>
<tr>
<td>Intention to use a wearable health device</td>
<td>0.258</td>
<td>0.478</td>
<td>0.542</td>
<td>0.620</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)N/A: not applicable.

Hypotheses Testing

The results obtained from testing H1, using data from both groups, show that the direct effect of trust propensity on trust is statistically insignificant at 0.11 (95% CI 0.048-0.177), as the CI of this direct effect includes 0 (Multimedia Appendix 3). In contrast, the indirect effect of organizational reputation is statistically significant at 0.22 (95% CI 0.143-0.309). We also split the data according to participants’ country of residence. This is to determine the differences in the mediating role of organizational reputation on the relationship between trust propensity and trust. The results show that organizational reputation has a statistically significant mediating role in the relationship between trust propensity and trust for both cohorts, albeit slightly lower in the case of the South African participants. These results support the hypothesis that the relationship between trust propensity and trust is mediated by organizational reputation.

The result of the regression analysis to test H2 is shown in Multimedia Appendix 4. The analysis reveals that there is no significant direct relationship between trust propensity and the intention to use a wearable health device with a direct effect of 0.001 (95% CI –0.056 to 0.058). When the indirect effect of trust is considered, the relationship between trust propensity and the intention to use a wearable health device becomes statistically significant at 0.26 (95% CI 0.172-0.349). Data from both countries were also analyzed separately. The results show that although trust plays a statistically significant mediating role in the relationship between trust propensity and the intention to use a wearable health device for the Kenyan participants, this is not the case for the South African cohorts. Hence, based on the combined results and the ones for Kenyan participants, the hypothesis that the relationship between trust propensity and the intention to use a wearable health device is mediated by trust is accepted.

Multimedia Appendix 5 illustrates the results of the mediation regression analysis to test H3 (the relationship between organizational reputation and the intention to use a wearable health device is mediated by trust). As shown in Multimedia Appendix 5, there is a statistically significant direct relationship between organizational reputation and the intention to use a wearable health device at 0.32 (95% CI 0.175-0.483). Although the mediating role of trust in the relationship between organizational reputation and the intention to use a wearable health device is statistically significant at 0.26 (95% CI 0.172-0.349), the relationship between organizational reputation and the intention to use a wearable health device is mediated by trust.
health device is also statistically significant at 0.35 (95% CI 0.218-0.494), this is not significantly different from the direct relationship between organizational reputation and the intention to use a wearable health device. A comparative analysis of data from both countries mirrors the combined result. Organizational reputation has a statistically significant direct relationship with the intention to use a wearable health device, with trust playing a slightly less indirect mediating role in the relationship for both cohorts. These results did not support H3. Thus, the hypothesis is rejected.

Multimedia Appendix 6 shows the results of the mediation regression analysis to test H4. The results show that there is no statistically significant relationship between perceived security and the intention to use a wearable health device with a direct effect of 0.07 (95% CI 0.000-0.147). This can be contrasted with the statistically significant indirect effect of trust at 0.36 (95% CI 0.255-0.461). A comparative analysis of data from both countries shows a similar pattern to the combined result. Thus, the results support our hypothesis that the relationship between perceived security and the intention to use a wearable health device is mediated by trust.

The results of the mediation regression analysis to test H5 are shown in Multimedia Appendix 7. Similar to the results of H4, there is no statistically significant relationship between perceived privacy and the intention to use a wearable health device with a direct effect of 0.07 (95% CI −0.009 to 0.150). The indirect effect of trust is however statistically significant at 0.42 (95% CI 0.282-0.557). The result of the comparative analysis of data from both countries is in line with the combined result. However, the result from Kenyan participants shows an indirect effect of 0.51 (95% CI 0.318-0.711). This is significantly higher than their South African counterparts where the indirect effect is 0.31 (95% CI 0.157-0.508). On the basis of the results, the hypothesis that the relationship between perceived privacy and the intention to use a wearable health device is mediated by trust is supported.

Table 3 provides a summary of the research hypotheses, the direct and indirect effects of the dependent and mediating variables for both countries and an individual country, and the hypotheses that are supported.

Table 3. Summary of research hypotheses findings.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Mediation path</th>
<th>Direct effect (C')</th>
<th>Indirect effect (ab)</th>
<th>Hypothesis supported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>All data^a</td>
<td>Kenya</td>
<td>All data^b</td>
</tr>
<tr>
<td>H1</td>
<td>Trust propensity - organizational reputation - trust</td>
<td>0.11</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>H2</td>
<td>Trust propensity - trust - intention to use a wearable health device</td>
<td>0.001</td>
<td>–0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>H3</td>
<td>Organizational reputation - trust - intention to use a wearable health device</td>
<td>0.32</td>
<td>0.73</td>
<td>0.31</td>
</tr>
<tr>
<td>H4</td>
<td>Perceived security - trust - intention to use a wearable health device</td>
<td>0.07</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>H5</td>
<td>Perceived privacy - trust - intention to use a wearable health device</td>
<td>0.07</td>
<td>0.04</td>
<td>0.09</td>
</tr>
</tbody>
</table>

^a All data: combined results.
^b SA: South Africa.

As stated in the Methods section, we used the Pillai trace to test for any statistically significant difference between the Kenyan and South African cohorts because of the slight violation of the Levene test of equality of variances by the variable, intention to use a wearable health device. The MANOVA result suggests that there are statistically significant differences between Kenyans and South Africans in the combined dependent variables (F_{6.22}=4.18; P=.001; Pillai trace=.1; partial eta squared=.01). We also considered the results of each dependent variable separately using a Bonferroni-adjusted alpha value of .008. There are statistically significant differences in the variables for both groups (Table 4). The only exception is in the organizational reputation variable, where there is no significant difference between the 2 cohorts. The mean ratings of the variables, as illustrated in Table 5, show that Kenyan participants’ ratings of the 6 variables are slightly higher than their South African counterparts. Table 5 also shows that despite the statistically significant differences between the 2 cohorts, the actual differences are less than 2 scales on average. The actual difference in the ratings for organizational reputation is less than 1 scale, which supports the results presented in Table 4.
### Table 4. Multivariate analysis of variance results showing differences in variables between groups.

<table>
<thead>
<tr>
<th>Variable</th>
<th>F value (df=1)</th>
<th>P value</th>
<th>Partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust propensity</td>
<td>9.72</td>
<td>.002&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.41</td>
</tr>
<tr>
<td>Perceived security</td>
<td>20.41</td>
<td>&lt;.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.82</td>
</tr>
<tr>
<td>Perceived privacy</td>
<td>13.47</td>
<td>&lt;.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.55</td>
</tr>
<tr>
<td>Trust</td>
<td>12.25</td>
<td>.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.51</td>
</tr>
<tr>
<td>Organizational reputation</td>
<td>6.35</td>
<td>.01&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.27</td>
</tr>
<tr>
<td>Intention to use a wearable health device</td>
<td>11.98</td>
<td>.001&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.49</td>
</tr>
</tbody>
</table>

<sup>a</sup>Statistically significant difference between groups.

<sup>b</sup>No significant difference between groups.

### Table 5. Participants’ mean score according to country of residence.

<table>
<thead>
<tr>
<th>Variable and country of residence</th>
<th>Mean</th>
<th>SE</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust propensity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>13.179</td>
<td>0.415</td>
<td>12.362-13.996</td>
</tr>
<tr>
<td>South Africa</td>
<td>11.496</td>
<td>0.345</td>
<td>10.816-12.177</td>
</tr>
<tr>
<td>Perceived security</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>12.653</td>
<td>0.358</td>
<td>11.947-13.358</td>
</tr>
<tr>
<td>South Africa</td>
<td>10.547</td>
<td>0.298</td>
<td>9.960-11.135</td>
</tr>
<tr>
<td>Perceived privacy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>11.947</td>
<td>0.388</td>
<td>11.183-12.712</td>
</tr>
<tr>
<td>South Africa</td>
<td>10.095</td>
<td>0.323</td>
<td>9.458-10.731</td>
</tr>
<tr>
<td>Trust</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>8.779</td>
<td>0.294</td>
<td>8.199-9.359</td>
</tr>
<tr>
<td>South Africa</td>
<td>7.438</td>
<td>0.245</td>
<td>6.955-7.921</td>
</tr>
<tr>
<td>Organizational reputation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>5.747</td>
<td>0.181</td>
<td>5.390-6.104</td>
</tr>
<tr>
<td>South Africa</td>
<td>5.153</td>
<td>0.151</td>
<td>4.856-5.451</td>
</tr>
<tr>
<td>Intention to use a wearable health device</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>5.716</td>
<td>0.222</td>
<td>5.278-6.154</td>
</tr>
<tr>
<td>South Africa</td>
<td>4.715</td>
<td>0.185</td>
<td>4.351-5.080</td>
</tr>
</tbody>
</table>

### Discussion

#### Principal Findings

The results obtained from this study showed that organizational reputation has a significant direct influence on the trust in and intention to use a wearable health device. Study participants’ country of residence did not change the direct influence of organizational reputation on the trust in and intention to use a wearable health device. Similarly, trust has a significant mediating effect on the intention to use a wearable health device even in the presence of perceived security and perceived privacy. The results from this empirical study of 232 participants (Kenya, n=137; South Africa, n=97) using mediation regression analyses support hypotheses H1, H2, H4, and H5. The one-way, between-group MANOVA test also supports H6.

As hypothesized, organizational reputation has a statistically significant mediating effect on the relationship between trust propensity and trust. This mediating effect was present when data from the 2 countries were analyzed together and when we analyzed the data according to the country of residence. Similarly, as predicted, trust has a statistically significant mediating effect on the relationship between trust propensity and intention to use a wearable health device. This is the case for the combined data and Kenyan participants’ data. However, trust did not have a significant mediating effect on the relationship between trust propensity and the intention to use a wearable health device for South African participants.
Results from our study show that trust propensity on its own does not necessarily lead to the intention to use a wearable health device. Previous studies, such as those by Zhou [36], Colesca [37], and Lee and Turban [38], found that people’s trust propensity influences their trust in and intention to adopt new technologies. Findings from our study are in line with previous studies in this regard. However, this study goes beyond previous studies by suggesting that organizational reputation significantly mediates the relationship between trust propensity and trust. Our research shows that people with high trust propensity will be more likely to trust a wearable health device when the device’s manufacturer has a good reputation.

The hypothesis that trust plays a significant role in the relationship between organizational reputation and the intention to use a wearable health device is supported by results from our study. Organizational reputation has a significant direct effect on the intention to use a wearable health device, with trust playing a lesser mediating role. Even when the country of residence is taken into account, organizational reputation on its own plays a significant direct role in the intention to use a wearable health device. Evidence from the literature suggests that good organizational reputation influences consumers’ trust and continued loyalty, irrespective of whether they live in a developed or developing country [47,50-52]. Thus, our study is in line with the ones developed (eg, America and Spain [50,52]) and developing (eg, Iran and Nigeria [47,51]) countries. Our study also confirms the value of good organizational reputation.

Results from our study also show that perceived security and perceived privacy on their own do not significantly influence consumers’ intention to use wearable health devices, rather these factors are mediated by trust. Previous studies, such as those by Gu et al [28], Arpaci [76], and Damghanian et al [77], investigated perceived security and perceived privacy in relation to trust, not the intention to use new technologies. Our study extends these studies by presenting empirical evidence that the presence of perceived security and perceived privacy on their own do not have a significant influence on consumers’ intention to use a wearable health device. Rather, trust plays a significant mediating role in the intention to use a wearable health device.

Our assumption that there is a significant difference between Kenyans and South Africans in their overall intention to use wearable health devices is supported by the study results. There are statistically significant differences between the 2 groups on the variables trust propensity, perceived security, perceived privacy, trust, and intention to use a wearable health device. However, there is no significant difference between the 2 cohorts on the variable organizational reputation. The MANOVA results also confirm those from the mediation regression analyses, where organizational reputation is shown to have a significant direct influence on the intention to use a wearable health device for the 2 cohorts.

Although we did not delve deeper into the reasons for the differences in the factors that influence Kenyans and South Africans in their intention to use wearable health devices, the results are in line with our expectation. For instance, Morawczynski and Miscione [78] found that Kenyans are more trusting of the provider of M-PESA (mobile money) service (ie, the organization) due to its proven track record. They are, however, less trusting of individual agents that often act as intermediaries between customers and the service provider. Similarly, a study by the South African Human Sciences Research Council [79] found that South Africans generally have low levels of trust. However, in spite of their low level of trust, South Africans tend to demonstrate an increased level of loyalty to brands with a good image [80]. Hence, results from our study confirm what has been reported in the literature about the influence of positive organizational reputation on Kenyan and South African consumers.

Study Contributions and Implications

The findings of this study made theoretical and methodological contributions to the field of uHealth. First, from a theoretical point of view, studies on the factors that influence the intention to use or adopt health technologies [28,30] did not consider the mediating effect of factors such as trust and organizational reputation. In contrast, we used a mediation regression analysis to provide deeper insight into the factors that mediate the relationship between a specific independent and dependent variable. Furthermore, our study is different from previous studies in that it is a cross-country study, whereas the previous studies are single-country studies. The cross-country nature of our study enriches the findings. Our study provides empirical evidence that organizational reputation has a significant mediating role in the relationship between trust propensity and trust. This study also confirms the fact that good organizational reputation has a positive influence on consumers’ trust in and intention to use wearable health technologies. Many of the studies on the relationship between organizational reputation and trust are predominantly about trust in web-based services [30,41,49,52]. Given the significant growth in the adoption of wearable health devices for monitoring PA, it is important to understand the influence that corporate image and trust have on consumers’ choice of wearable health devices.

Second, from a methodological point of view, our study uses an emerging approach for the mediation regression analysis, the PROCESS macro by Hayes (see the Methods section for some of the benefits of PROCESS) [74]. Our study demonstrates the effectiveness of the PROCESS macro by Hayes [74] and adds to the growing number of studies, such as those by Naidoo [81], Zhang et al [82], Huang et al [83], Barboza and Siller [84], Ahmed et al [85], and Supakong and Jarunratanakul [86], that uses the method.

From a practical perspective, our study has implications for the manufacturers of wearable health devices. Results from the study demonstrate the significant influence of organizational reputation and trust on consumers’ intention to use wearable health devices. In the absence of trust and good organizational reputation, factors such as trust propensity, perceived security, and perceived privacy may not necessarily drive consumers to adopt wearable health devices. A good reputation is an organizational asset [43-45]. Manufacturers of wearable health devices should capitalize on this asset to attract more potential adopters of wearable health devices.
Limitations and Future Research

Although this study made theoretical, methodological, and practical contributions, it has limitations. For instance, we did not consider the moderating role of age on the factors that influence the intention to use wearable health devices. Furthermore, we did not explore the factors that could explain the differences in the 2 cohorts’ intention to use a wearable health device. Future studies can extend the research model by specifically investigating the roles of age and previous experience in the use of a wearable health device by using these variables as moderators to the factors that influence the intention to use a wearable health device. Such a study could also explore the potential effect of fitness tracker purchasing habits on the factors that influence the intention to use wearable health devices.

Conclusions

In this study, we investigated the factors that influence consumers’ intention to use wearable health devices. More specifically, we considered the mediating roles of organizational reputation and trust in the intention to use a wearable health device. We collected data from 232 participants resident in Kenya and South Africa through a questionnaire implemented on Google Forms. A mediation regression analysis and MANOVA tests were used to analyze data. The study provides empirical evidence that organizational reputation has a significant mediating effect on the relationship between trust propensity and trust. In addition, we demonstrated that even when the country of residence is taken into account, organizational reputation on its own significantly influences consumers’ intention to use a wearable health device. Another important finding from the research is that perceived security and perceived privacy are not sufficient to motivate the use of wearable health devices. Trust is an important factor that drives this intention. The study provides deeper insight into the factors that influence the intention to use a wearable health device by investigating the factors that mediate this intention, as opposed to linear relationships between the factors. Without good organizational reputation and trust, factors such as trust propensity, perceived security, and perceived privacy have little influence on consumers’ intention to use a wearable health device.

Acknowledgments

The authors would like to thank the study participants. They also express their appreciation to Prof Rennie Naidoo, Prof Hanlie Smuts, and the reviewers for their constructive criticisms of the manuscript. They acknowledge the Department of Informatics and the University of Pretoria for providing the funding for the article processing fee.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research model.
[ PNG File, 15 KB - mhealth_v8i6e16721_app1.png ]

Multimedia Appendix 2
Research instrument.
[ PDF File (Adobe PDF File), 89 KB - mhealth_v8i6e16721_app2.pdf ]

Multimedia Appendix 3
Mediating role of organizational reputation in the relationship between trust propensity and trust, with regression coefficients, indirect effects, and bootstrapped CI.
[ PNG File, 19 KB - mhealth_v8i6e16721_app3.png ]

Multimedia Appendix 4
Mediating role of trust in the relationship between trust propensity and intention to use wearable health devices, with regression coefficients, indirect effects and bootstrapped CI.
[ PNG File, 19 KB - mhealth_v8i6e16721_app4.png ]

Multimedia Appendix 5
Mediating role of trust in the relationship between organization reputation and intention to use wearable health devices, with regression coefficients, indirect effects, and bootstrapped CI.
[ PNG File, 20 KB - mhealth_v8i6e16721_app5.png ]

Multimedia Appendix 6
Mediating role of trust in the relationship between perceived security and intention to use wearable health devices, with regression coefficients, indirect effects and bootstrapped CI.

Multimedia Appendix 7
Mediating role of trust in the relationship between perceived privacy and intention to use wearable health devices, with regression coefficients, indirect effects and bootstrapped CI.

References


Abbreviations

- CWHD: consumer wearable health device
- eCommerce: electronic commerce
- eGovernment: electronic government
- MANOVA: multivariate analysis of variance
- NFC: near-field communication
- PA: physical activity
- uHealth: ubiquitous health

http://mhealth.jmir.org/2020/6/e16721/ JMIR Mhealth Uhealth 2020 | vol. 8 | iss. 6 | e16721 | p.656 (page number not for citation purposes)
The Mediating Role of Organizational Reputation and Trust in the Intention to Use Wearable Health Devices: Cross-Country Study

Adebesin F, Mwalugha R

Please cite as:
Adebesin F, Mwalugha R
The Mediating Role of Organizational Reputation and Trust in the Intention to Use Wearable Health Devices: Cross-Country Study
JMIR Mhealth Uhealth 2020;8(6):e16721
URL: http://mhealth.jmir.org/2020/6/e16721/
doi:10.2196/16721
PMID:32348260

©Funmi Adebesin, Revingstone Mwalugha. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 09.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Continuous Measurement of Reconnaissance Marines in Training With Custom Smartphone App and Watch: Observational Cohort Study

Leslie Saxon¹, MD; Brooks DiPaula¹, BA; Glenn R Fox¹,², PhD; Rebecca Ebert¹, BSc; Josiah Duhaime³, BSc; Luciano Nocera⁴, PhD; Luan Tran⁴, BSc; Mona Sobhani¹, PhD

¹University of Southern California, Center for Body Computing, Keck School of Medicine, Playa Vista, CA, United States
²University of Southern California, Marshall School of Business, Los Angeles, CA, United States
³United States Marine Corps, Reconnaissance Training Company, Camp Pendleton, CA, United States
⁴University of Southern California, Department of Computer Science, Viterbi School of Engineering, Los Angeles, CA, United States

Corresponding Author:
Leslie Saxon, MD
University of Southern California
Center for Body Computing
Keck School of Medicine
12015 East Waterfront Drive
Playa Vista, CA, 90094
United States
Phone: 1 310 448 5373
Email: saxon@usc.edu

Abstract

Background: Specialized training for elite US military units is associated with high attrition due to intense psychological and physical demands. The need to graduate more service members without degrading performance standards necessitates the identification of factors to predict success or failure in targeted training interventions.

Objective: The aim of this study was to continuously quantify the mental and physical status of trainees of an elite military unit to identify novel predictors of success in training.

Methods: A total of 3 consecutive classes of a specialized training course were provided with an Apple iPhone, Watch, and specially designed mobile app. Baseline personality assessments and continuous daily measures of mental status, physical pain, heart rate, activity, sleep, hydration, and nutrition were collected from the app and Watch data.

Results: A total of 115 trainees enrolled and completed the study (100% male; age: mean 22 years, SD 4 years) and 64 (55.7%) successfully graduated. Most training withdrawals (27/115, 23.5%) occurred by day 7 (mean 5.5 days, SD 3.4 days; range 1-22 days). Extraversion, positive affect personality traits, and daily psychological profiles were associated with course completion; key psychological factors could predict withdrawals 1-2 days in advance \( (P=.009) \).

Conclusions: Gathering accurate and continuous mental and physical status data during elite military training is possible with early predictors of withdrawal providing an opportunity for intervention.

(JMIR Mhealth Uhealth 2020;8(6):e14116) doi:10.2196/14116

KEYWORDS
military; marines; wearable devices; wearable technology; smartphone; mobile app

Introduction

Background
Owing to the nature of the US conflicts in Iraq and Afghanistan, which span nearly two decades, the US military comprised many experienced warfighters who are highly trained and specialized [1,2]. Service members volunteer for training and selection to serve in elite and specialized US military units. One such group, Reconnaissance (Recon) Marines, undergoes intense and specialized psychological and physical training that is associated with high attrition rates, most of which occurs in the initial phase of training, and as a result, only approximately half are
able to complete the nearly 90-day initial training course [3]. There is widespread acknowledgment that those who complete the course are highly adept warfighters, many of whom will serve for one to two decades.

Owing to continued global threats, there is a need to graduate more highly trained Recon Marines, without degrading performance standards [4,5]. Yet, as Recon training is exceedingly rigorous, the majority of the training failures are trainee-initiated, commonly referred to as Drop on Request (DOR) [3]. It is not known which mental and physical factors are predictors of success or failure in training. Some previous studies have used existing military assessments to investigate factors that best predict success or failure [3], whereas others have incorporated wearable sensors to collect and analyze physiologic signals [6,7] or analyzed personality and psychological variables as predictors [8-12]. However, none of these studies have analyzed existing military assessments along with continuously collected physiologic, personality, and psychological data to predict success or failure in military trainings. Furthermore, no studies have collected longitudinal and continuous physiological signals, and few studies have been conducted in naturalistic military training scenarios without disrupting standard procedures.

**Objectives**

To help determine novel predictors of success or failure in training, we collected continuous mental and physical status of Recon trainees to build comprehensive models capable of identifying the collective load of Recon training and the thresholds at which trainees are most likely to remove themselves from the course. In doing so, novel insight into the mental and physical processes associated with these factors can inform the theories of human performance and improve desired outcomes in elite military training.

The University of Southern California, Center for Body Computing, a digital health and human performance research center, partnered with Reconnaissance Training Company leadership at Camp Pendleton, California, in 2016 to design and implement a human performance study aimed at obtaining continuous assessments of Recon trainees, using connected technology to garner insights and identify novel predictors of training completion. Over the course of 36 months, requirements were identified that included the need to collect (1) personality profiles, as prior military studies had identified personality traits such as grit to be predictive of training success [4,13,14]; (2) physiologic data including heart rate and work output (measured via accelerometer and photoplethysmography activity and calorie expenditure) during training exercises taking place in both land and water; (3) daily assessment of sleep, nutrition, and hydration; and (4) daily assessment of mental and physical pain and attitudes regarding successful completion of training. Our collective goal was to understand whether this new information could have additive value to standard measures of performance already being collected as well as be used to confirm or refute subjective instructor evaluation of trainee status. Another critical concern was to develop a study protocol that collected these continuous measures without disrupting the culture and cadence of training. We also identified the need to develop specialized data collection processes, sensors, and software that would engage and achieve adherence from the trainees throughout the course.

**Methods**

**Study Subjects**

The 3 successive classes of Marines and Sailors entering the 25-day Basic Reconnaissance Primer Course (BRPC) of the Reconnaissance Training Company at Camp Pendleton, California, were offered study enrollment (April 26, 2018, to August 17, 2018) on the first day of orientation (Figure 1). A total of 60.5% (121/200) of the trainees consented to participate. Study personnel explained that the study was completely voluntary and that the study data would not be shared with course instructors. Of the 121 who consented, a total of 6 (5.0%) trainees dropped out of the study, and we did not collect data on their reasons for withdrawal. The study was approved by the Institutional Review Boards of the University of Southern California and Marine Corps Training and Education Command (University of Southern California IRB HS-17-00729 and Training and Education Command Human Subject Research Approval DoDI 3216.16D). Sample size was determined by the number of trainees willing to participate in the study. Informed consent was obtained using the mobile app. Baseline demographics of the study subjects are summarized in Table 1.
Table 1. Baseline demographics of study subjects.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>22 (4)</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>175 (8)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>77 (9)</td>
</tr>
<tr>
<td>Participated in high school sports (N=104), n (%)</td>
<td>79 (76.0)</td>
</tr>
<tr>
<td>Marital status: single (N=115), n (%)</td>
<td>107 (93.0)</td>
</tr>
<tr>
<td>Time in military (months), mean (SD)</td>
<td>23 (23)</td>
</tr>
<tr>
<td>Military rank (N=115), n (%)</td>
<td></td>
</tr>
<tr>
<td>Marine officer</td>
<td>4 (3.5)</td>
</tr>
<tr>
<td>Marine enlisted</td>
<td>105 (91.3)</td>
</tr>
<tr>
<td>Navy enlisted</td>
<td>12 (10.4)</td>
</tr>
</tbody>
</table>

Basic Reconnaissance Primer Course Measures

The 25-day BRPC is the initial phase of Reconnaissance training that must be successfully completed to advance to the 60-day Basic Reconnaissance Course (BRC), which is required for graduation. Marines and Sailors are accepted into the BRPC from other marine operational specialties or after enlistment and completion of entry-level infantry training.

The BRPC consists of structured daily training on land and water that embeds the cognitive and physical assessments that trainees have to complete, benchmarked to time or quality standards. The aquatics training includes various high-intensity underwater and treading drills that impose intense psychological and physical stressors. This course selectively screens and prepares trainees to acquire the mental and physical skills taught in BRC, such as land navigation, coastal piloting, and radio communications, which render them able to serve as a Reconnaissance Marine within the Fleet Marine Forces (Figure 2).
Figure 2. Photographs of trainees engaged in land exercises (a) and water exercises (b).

Reasons for Training Failures
Trainees can drop out of training by choice on request (DOR), and medical personnel can remove trainees from the course if they have an illness or injury (medical) or instructors determine that trainees pose a safety hazard to themselves or other trainees (safety) or fail to meet one or more of the course training standards (performance). Trainees that DOR are not permitted to repeat the training course and are then assigned back to their prior unit or are assigned to a different training school to obtain another military occupational specialty. Those who are excluded from the course for medical, safety, or performance reasons may repeat the course at the next scheduled course initiation.

Study Software and Hardware
A specially designed software app, used for subject consent and data collection, was designed on Apple’s ResearchKit platform (Apple Inc and Thread Research) for the purposes of this study (Figure 3) [15-17]. At the time of consent, iPhones and Apple Watches were distributed to study subjects, and they received a structured tutorial on how to use the app and Watch and were instructed to wear the Watch for collection all day and all night. We expected that attrition due to compliance and user error could be a substantial hurdle to gathering complete datasets, as the paradigm required subjects to use the Apple Watch to log in their workouts via workout mode before each activity, which allowed for stronger data collection but caused a need for the battery to be recharged every few hours. To increase compliance, study personnel supported Watch and iPhone battery charging during training days at scheduled breaks and redistributed them to trainees up to 4 times during a 24-hour cycle to increase compliance. Study personnel also worked with instructors to ensure that all subjects had their devices before beginning the class. During high-intensity training drills, such as land-based or aquatic training, Watches were put in the workout mode to collect data with a higher sampling rate (sampling heart rate continuously every 5 seconds vs the standard sampling frequency of every 5 min) [18,19].

Figure 3. Sample study app screenshots: (a) initial study screen, (b) prestudy survey screen, and (c) daily survey.
Study Measures

Table 2 lists all study metrics that were collected. A range of validated surveys were administered at the beginning of the study that assessed various personal characteristics, such as personality type, emotional processing, outlooks on life, and mindfulness. To assess personality, we used the Big Five Inventory (BFI), which is a well-validated personality construct that has been previously used in military human performance and leadership research [8,11,20-23]. We also included the Psychopathic Personality Inventory-Revised to capture additional personality traits [24]. As positive affect has been associated with sport performance [25] and resilience [26], we included the Positive and Negative Affect Scale (PANAS) [27]. To assess outlook on life and mindfulness, we used the Satisfaction with Life Scale (SWLS) [28] and Five Facet Mindfulness Questionnaire (FFMQ) [29], respectively. As grit and resilience have been identified as important traits for success in intense military trainings from previous research [12,13,15], we included the Grit Scale [12,13] and the Ego Resilience Scale (ERS) [30].

Table 2. Study metrics by collection method.

<table>
<thead>
<tr>
<th>Method and metric</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>iPhone app</strong></td>
<td></td>
</tr>
<tr>
<td>Consent</td>
<td></td>
</tr>
<tr>
<td>Questionnaires: prestudy</td>
<td>• Demographic survey</td>
</tr>
<tr>
<td></td>
<td>• Big Five Inventory [20-22]</td>
</tr>
<tr>
<td></td>
<td>• Psychopathic Personality Inventory-Revised [24]</td>
</tr>
<tr>
<td></td>
<td>• Positive and Negative Affect Schedule [27]</td>
</tr>
<tr>
<td>Daily questionnaires (Multimedia Appendix 1)</td>
<td>• Satisfaction with Life Scale [28]</td>
</tr>
<tr>
<td></td>
<td>• Five Facet Mindfulness Questionnaire [29]</td>
</tr>
<tr>
<td></td>
<td>• Grit Scale [12,13]</td>
</tr>
<tr>
<td></td>
<td>• The Ego Resilience Scale [30]</td>
</tr>
<tr>
<td><strong>Apple Watch</strong></td>
<td></td>
</tr>
<tr>
<td>Heart rate (beats per minute)</td>
<td></td>
</tr>
<tr>
<td>Daily activity</td>
<td>• Mental and physical pain scale (1-5)</td>
</tr>
<tr>
<td></td>
<td>• Sleep, hydration, and nutrition (self-report)</td>
</tr>
<tr>
<td></td>
<td>• Confidence in instructors and graduation (4 questions)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Watch metrics were visually inspected for outliers using histograms of all raw scores for each subject across all workouts. Values of zero (<0.0001% of all values) were removed from the analyses because they could skew data and likely resulted from the Watch not being worn. The R outlier package was used to identify if zero values were sufficiently outside the distribution. Wilcoxon rank sum tests were used to assess differences between groups on Apple Watch metrics including heart rate, daily steps, active energy, and resting energy as well as daily survey assessments. To investigate the factors that may specifically contribute to voluntary course withdrawal (DORs), we compared the DOR and Pass groups directly for certain metrics, including baseline personality assessments and daily survey assessments. A random forest classifier was also used to classify course completion status from baseline personality surveys [35,36]. Wilcoxon rank sum tests were used to assess differences between DOR and Pass groups’ mean ratings for daily survey assessments on a day-to-day basis.

Statistical Plan

Statistical analyses were performed using R Statistical Software (Bell Laboratories) [31]. Percentages were calculated for course completion categories. For the DOR withdrawal requests, the time of day and nearest training event were analyzed for patterns. To examine whether there was an association between course completion categorical variable and students’ age, weight, height, and years in the Marine Corps, one-way analysis of variance was performed. A chi-square test of goodness-of-fit was performed to assess whether there was an equal distribution of rates of marital status, military rank, number of high school sports played, and number of previous BRPC attempts across course completion categories [32]. Normality of data was assessed using Kolmogorov-Smirnov tests [33]. The differences between course completion categories on physical test standard assessments were assessed using Wilcoxon rank sum tests [34], with significance set at \( P < 0.05 \), two-tailed. The 3-mile hike was performed by only 2 cohorts, and the distribution was as follows: Pass (ie, successful course completion), n=46; DOR, n=6; medical, n=6; safety, n=6; and performance, n=4. Descriptive statistics for the Apple Watch daily metrics were calculated for each individual using the number of days in the course. The
Results

Course Completion

The summary data for the course completion of all 3 BRPC study subjects are shown in Multimedia Appendix 2. Of the 115 subjects who completed the study, a total of 64 (55.7%) trainees successfully completed the course, and the reasons for failure to complete were DORs (27/115, 23.5%), medical (10/115, 8.7%), safety (8/115, 7.0%), and performance (6/115, 5.2%). The majority of training failures were trainee-initiated DORs, and these occurred, with one exception, by day 7 (mean 5.5 days, SD 4 days; range 1-22 days). There were no significant differences in course completion rates between successive classes ($X^2=15.1; P>.05$).

The timing and context of DORs fell into a consistent pattern, and 93% of drops occurred before an impending aquatic event or in the training pool. The DORs that did not occur before an impending aquatic event were observed at the beginning of the training day in a waiting area outside of the training schoolhouse. There were no significant differences between course completion groups with respect to age, height, weight, history of high school sports participation, marital status, years in the Marine Corps, Marine rank, or previous BRPC attempts ($P>.05$).

Course Training Standards and Course Completion

Figure 4 shows the course completion results of all study subjects according to the physical test standard assessments of sit-ups, pull-ups, and a timed 3-mile run. The median number of sit-ups performed was not significantly different between the Marines who completed the course and each category of failure ($P>.05$). The number of pull-ups was significantly lower in the DOR group ($P=.049$) and in the group that failed for medical reasons ($P<.05$) compared with the group who completed the course. The time on the 3-mile run standard trended toward a significant difference between the group that requested a DOR compared with those that passed ($P=.05$). In addition, there was no significant difference between the median 3-mile hike time and the reason for training withdrawal ($P>.05$).

Figure 4. Course completion results for test standard assessments of (a) sit-ups; (b) pull-ups, timed; and (c) 3-mile run.

Physiologic Apple Watch Data and Course Completion

Table 3 provides a summary of 24-hour heart rate data and day and night heart rate data for the subject cohort over the 25-day course, according to course completion categories. Outlier analyses revealed no outliers in the data. The range values displayed in the table indicate that at least one subject hit a minimum heart rate of 30 beats per minute (BPM) or a maximum heart rate of 210 BPM at least once during the training period, as the Apple Watch heart rate sensor supports a range of 30-210 BPM [37]. Daily heart rates up to 210 BPM were recorded in most subjects during training days, and mean heart rates at night (between the hours of 8 PM and 4 AM), corresponding to sleep hours, were consistently 50% of mean rates during daytime hours. There were no significant associations between mean, range, daytime or nighttime heart rates in subjects who completed versus those who did not complete training, or in categories of unsuccessful trainees ($P>.05$).

Table 4 compares the mean daily step count of all study subjects according to the completion category. Subjects who passed the course had higher mean daily step counts than those who did not ($P<.05$), and this was significant for all categories of noncompletion (DOR, $P<.005$; medical, $P<.005$; safety, $P<.05$; performance, $P<.05$). There were no significant differences in daily step counts between noncompletion categories.

Table 5 compares the mean active and resting energy expenditure (calorie) of the entire cohort according to the completion category. Subjects who passed the course had higher active energy expenditure than those who did not ($P<.05$), and this was significant for all categories except performance noncompletion (DOR, $P<.001$; medical, $P<.05$; safety, $P<.05$; performance, $P<.06$). There were no significant differences in daily step counts between noncompletion categories. Resting energy expenditure did not differ between those completing the course and completion failures, with the exception of DOR failures ($P<.05$).
Table 3. Mean heart rate (beats per minute) per group (course completion and all noncompletion categories) compared by time of day.

<table>
<thead>
<tr>
<th>Group</th>
<th>24 hour</th>
<th>Day</th>
<th>Night</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Pass (BPM&lt;sup&gt;b&lt;/sup&gt;)</td>
<td>110 (30)</td>
<td>112 (29)</td>
<td>64 (16)</td>
</tr>
<tr>
<td>Drop on request (BPM)</td>
<td>106 (31)</td>
<td>111 (28)</td>
<td>64 (19)</td>
</tr>
<tr>
<td>Medical (BPM)</td>
<td>116 (29)</td>
<td>118 (28)</td>
<td>73 (19)</td>
</tr>
<tr>
<td>Safety (BPM)</td>
<td>111 (29)</td>
<td>113 (27)</td>
<td>69 (10)</td>
</tr>
<tr>
<td>Performance (BPM)</td>
<td>118 (28)</td>
<td>119 (27)</td>
<td>70 (13)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Apple Watch heart rate sensor supports a range of 30-210 beats per minute [37].
<sup>b</sup> BPM: beats per minute.

Table 4. Mean activity (daily steps) per group (course completion and all noncompletion categories).

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass (daily steps)</td>
<td>10,537 (3195)</td>
<td>500-15,311</td>
</tr>
<tr>
<td>Drop on request (daily steps)</td>
<td>4491 (4541)</td>
<td>239-15,057</td>
</tr>
<tr>
<td>Medical (daily steps)</td>
<td>4585 (4899)</td>
<td>456-15,537</td>
</tr>
<tr>
<td>Safety (daily steps)</td>
<td>6569 (4677)</td>
<td>780-11,787</td>
</tr>
<tr>
<td>Performance (daily steps)</td>
<td>7660 (3395)</td>
<td>2734-11,674</td>
</tr>
</tbody>
</table>

Table 5. Mean energy expenditure (calories) per group (course completion and all noncompletion categories) compared by activity level.

<table>
<thead>
<tr>
<th>Group</th>
<th>Active&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Resting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Range</td>
</tr>
<tr>
<td>Pass (calories)</td>
<td>1837 (1231)</td>
<td>1-9133</td>
</tr>
<tr>
<td>Drop on request (calories)</td>
<td>1329 (1266)</td>
<td>1-4911</td>
</tr>
<tr>
<td>Medical (calories)</td>
<td>961 (1294)</td>
<td>2-6021</td>
</tr>
<tr>
<td>Safety (calories)</td>
<td>1271 (1445)</td>
<td>1-5746</td>
</tr>
<tr>
<td>Performance (calories)</td>
<td>1450 (971)</td>
<td>11-3667</td>
</tr>
</tbody>
</table>

<sup>a</sup> Apple Watch uses logged workouts to calculate active energy expenditure.

Survey Responses and Course Completion

The compliance of the baseline personality surveys was as follows: (1) BFI: 94.0% (108/115), (2) PANAS: 93.0% (107/115), (3) PPL: 76.5% (88/115), (4) SWLS: 92.2% (106/115), (5) ERS: 90.4% (104/115), (6) Grit: 89.6% (102/115), and (7) FFMQ: 83.5% (96/115). The baseline personality traits that were most associated with course completion include extraversion and positive affect.

Subjects who had DOR reasons for course failure versus those completing the course demonstrated an increase in emotional \(P<.03\) or physical pain \(P<.002\) and a degradation of confidence \(P=.05\), on a 1 to 5 grading scale, up to 3 days before withdrawal from the course.

There was no difference in self-reported feelings related to instructor support, adequacy of hydration, nutrition, or sleep duration between subjects completing or not completing the course. All trainees thought about quitting the course on at least one daily survey.

Discussion

Principal Findings

As the need for a modernized military grows, there is also a need to understand how best to involve larger numbers of warfighters in higher-level training and operations without degrading training and performance standards [2,4]. Here, we validated a model and platform yielding consistent and novel insights into successful warfighter training, the results of which can scale into other elite training and performance domains. We found novel physiological and psychological measurements capable of predicting success and failure in this difficult training course. Namely, the amount of physical energy expended early in training as well as positive affect and extraversion independently predict a lower probability of voluntarily leaving the course. These efforts represent a new model for research on military training and, perhaps most importantly, a validation of this approach for yielding insight past intuition regarding what constitutes high performance and fortitude in the face of training-related stressors.
Previous studies of elite military training have focused on psychological and physiological factors. In a previous study of Navy Sea, Air, and Land (SEAL) warfighters, compared with the general male population, SEALs scored lower in neuroticism and agreeableness and higher in extraversion and conscientiousness [38]. In a highly comprehensive review of factors pertinent for graduation from Basic Underwater Demolition/SEAL (BUD/S) training, Taylor et al. [5] reviewed 13 studies of BUD/S candidates’ success rates. Across the studies reviewed, general themes revealed a range of physical fitness factors associated with graduation from BUD/S, although the results were not consistent across studies. Psychologically, emotional stability, adjustment, and likeability were implicated in individual reports but were also not consistently found to be predictive, highlighting the need for further inquiry that combines psychological and physiological measurements using modern devices. Prior studies have also identified that greater physical fitness predicts resilience in stressful environments [39] and that mental skills training can help ameliorate the deleterious effects of intense deployments on both mental state and stress hormones [40], together indicating the key interplay and joining of mental and physical functions. Although this study does not investigate all the potential cognitive and physiological interactions, our approach lays groundwork to bridge and unify these existing studies into natural settings where modern wearable technology can be leveraged in combination with physical and mental factors.

### Novel Insights From Direct and Indirect Measures of Trainee Physical Status

On the basis of this prospective study, we see evidence that the standard physical fitness tests and physiological measurements provide limited information regarding a Marine’s or Sailor’s likelihood of voluntarily dropping from training. Of all 4 land training standard assessments, only pull-ups were predictive of an increased probability of a DOR. Similarly, heart rate data was not a good predictor, perhaps reflecting the fact that trainees were young and healthy, and there was limited variability in the range of heart health among the sample, suggesting that their cardiovascular health was adequate to complete training.

The mean step counts of those who successfully completed the course were significantly higher than those who failed to complete the course, most of whom were removed from the course by training day 7. Not all training days involve physical training; subjects spent training days in a classroom environment as well, hence the wide range in mean daily step counts. Still, mean daily step counts correlate to approximately 2 to 3 miles daily in those who withdrew versus a mean of over 5 miles daily in those who completed the course. This signifies that most of the work output of the training course occurred after trainees withdrew, supporting the notion that they did not withdraw for physical reasons, and this may hint at broader clues as to the importance of motivation and personal narrative in determining success.

Indirect measures of energy expenditure, such as active energy, also indicate that those who successfully completed the course expended higher activity over time, versus those who were unsuccessful at completing the course. This was true for all categories of unsuccessful course completion except performance withdrawals. Those who were unsuccessful at completing the course due to performance withdrawal also tended toward higher step counts and were dropped due to failure to complete cognitive and physical tasks. This metric is also important because it encompasses and can measure training work done in water, which is not accounted for in step counts. The energy expenditure data are also revealing in that they provide the first comprehensive data on the total mean daily energy expenditure of trainees. Although the range is considerable, the observation that in the trainees who successfully completed the course approximately 3300 calories/day are needed to maintain caloric equilibrium is important knowledge in terms of nutritional planning and avoidance of weight loss [30,31]. These interfaces between psychological variables and energy expenditure stand to contribute greatly to models of performance, as the data collected over time can be used to design improvements to, and efficiency of, training schedule and requirements.

Although previous studies have examined the existing military training assessments, such as physical fitness scores [3], for the BRC, none have examined the individual tests that comprise the score (eg, pull-ups, sit-ups, hikes, and run) and none have examined BRPC. Furthermore, no other studies have collected longitudinal and continuous physiologic data and other wearable data from trainees during the entire duration of a training course.

### Personality and Emotional Status Assessment

We found that higher levels of self-reported positive affect and extraversion were significantly associated with successful course completion. Moreover, although all trainees entertained thoughts of quitting, those that did voluntarily or involuntarily withdraw were more introverted and had less demonstrative personality types than those that successfully completed the course. In contrast to previous research performed in military academies or in other military training [4,12-14], subject scores on the Grit Scale did not predict success or failure. However, the findings of this study are in line with other research that indicates a positive relationship between positive affect and psychological resilience in the US Military [41], sport performance [25,42], coping strategies in competitive athletes [43], and task performance [44]. Previous research also demonstrates a relationship between extraversion and military leadership [9] as well as sport performance [45,46]. A more detailed understanding of these differences is enabled by the daily survey data showing that students who DOR’d had a significant degradation in their self-reported emotional and physical pain and confidence scores, compared with those who successfully completed the course 1 to 2 days before withdrawal. This underscores the importance of daily and individualized assessment to gain fidelity into understanding exactly why or when a trainee may decide to quit. This new understanding points to a clear target for early intervention. Studies have shown that positive affect [47,48] and other mental skills training interventions [40,49,50] are feasible and effective; thus, future studies should determine if more positive affect and optimistic traits can be cultivated in these trainees to improve course completion rates and resiliency. As mentioned above, the ability to gauge in depth which training element (eg, aquatic events...
and overland hikes) responsible for the greatest amount of trainee doubt and anxiety offers the ability to direct resources toward building readiness for and confidence in completion of those training tasks.

Validation of a New Model for Military Training Research

A unique aspect of this study is that it demonstrates the feasibility of a fully comprehensive and continuously collected model of human performance research in a naturalistic military training environment. We demonstrated the ability to collect accurate and continuous physical performance and psychological data throughout a 25-day course that takes place on land and in water. The success of the model was validated across the 3 successive training classes that we enrolled. The technology platform, Apple’s ResearchKit, HealthKit, and CareKit software stack coupled with the hardware components consisting of the iPhone and Watch provide important efficiencies and will further improve future studies through the integration of additional sensors and provide novel methods of data visualization in future apps [15-17]. The Apple technology stack was chosen for a few reasons. One reason is that biometric tracking on the Watch has been shown to be accurate and validated [18,19]. Another reason is that the cybersecurity and data privacy controls are state-of-the-art. In addition, we were able to leverage Apple’s ResearchKit to enroll subjects in the study through digital consent within the study app. Moreover, the devices were chosen because of subjects’ affinity for and familiarity with Apple products. The ResearchKit and HealthKit also provided the opportunity to bring in other survey data from the app and integrate Apple Watch and other data. Our app design process was collaborative and iterative. We regularly met with our software designers and the Reconnaissance Marines instructors as well as our research team to design and test versions of the app. The ResearchKit app enabled us to conduct a research study on an entirely digital platform, including the ability to consent subjects on their iPhones, and to custom design a Reconnaissance Marine app that served as the interface to the subject for collection of daily surveys.

Apple Watch data were collected using HealthKit, which provides the ability to collect Watch data and integrate it into the dataset collected on ResearchKit. In the future, we will leverage HealthKit to integrate other connected sensors such as weight, oxygen saturation, or temperature. The accuracy of the direct measures, such as heart rate, collected from the Watch using photoplethysmography, has been demonstrated by others to have a sensitivity and specificity of 98% and 90%, respectively [51]. As data in this study were continuously collected over the entirety of training days, we were also able to internally validate the heart rate measures by comparing daytime with nighttime readings, which were consistent with typical circadian variability of heart rates observed in a young, healthy male population [52]. The ability to perform research studies in a continuous model of data collection evolves the research model to one that is much more comprehensive and efficient than traditional models that focus on data collection during a single discrete study interval and are performed under more artificial and less naturalistic conditions than being embedded within the actual training environment being studied.

It is particularly important to collect data for military populations while they perform their normal functions, which often include stressful conditions that are not easy to reproduce in a laboratory setting.

Other benefits of this model include the ability to more quickly and efficiently translate study findings into interventions that can be tested in the subsequent classes. For example, our finding that declines in both daily self-reported emotional and physical pain scores and confidence in graduating preceded DORs from training provide an opportunity to target and test interventions. In addition, the finding that most DORs occurred in relation to an aquatic training event suggests that these events are causing significant doubt, distress, and lack of confidence. The finding that individuals who completed the course manifest more positive affect suggests that an effective intervention should instill these traits and behaviors. Ideally, an intervention targeted at cultivating confidence for these activities will heighten the probability of successful completion without changing the rigorous nature of the training itself. This could include specially designed educational or motivational video content, which is delivered to the trainee manifesting these thoughts and feelings, through the next version of the study software app. Similarly, advances in technologies, such as the real-time electrocardiogram (ECG) capability on the Apple 4 Series Watch, can be easily integrated into the study and provide another opportunity to validate the heart rate data. These ECG and heart rate capture capabilities may provide a real-time safety component for these trainees, while undergoing intense physical training. For instance, if a trainee sustained a very elevated heart rate, this could be validated by obtaining an ECG from the Watch and displayed to instructors in real time, using a CareKit-enabled data dashboard, which could then allow for timely medical assessment of that individual.

Future Studies

We have designed and implemented a flexible and continuous research model for military training that can easily accommodate best-in-breed sensors, allow for rapid software integration to more efficiently assess and test novel training interventions, and increase the course completion rate and quality of new Reconnaissance Marines and other military equivalents. This model can also provide important efficiencies and data flow that will help assess the medical well-being of trainees. We also found that despite the physical rigor involved in Reconnaissance Marine training, most of the attrition is due to mental deterioration rather than physical deterioration, and students who DOR often do so outside of the deepest physical stressors, that is, while on the pool deck as opposed to actually being in the pool. These signs of deterioration can be identified before trainees withdraw or are withdrawn from training, and interventions that target the reversal of these feelings as well as help to gain a deeper understanding of how to determine selection criteria can be designed.

Limitations

We did not collect data on the entirety of the courses offered by the Reconnaissance Training Company, namely, the follow-on 60-day BRC that follows the 25-day BRPC, which limits our ability to draw conclusions about fully graduating
into the Reconnaissance military occupational specialty. We also did not collect any data on weekends, mostly due to initial study design limitations with maintaining the study equipment inventory of iPhones, Watches, and data plans, which can be addressed through improved battery life and study design, leading to a model capable of measuring students during the entirety of trainee experience, including time off and personal time.

Daily questions (Multimedia Appendix 1) inquiring about substance use were only included in the second and third cohorts and inquired about nicotine and alcohol but not caffeine and other substances known to affect performance. In addition, the response scales were not standardized. For these reasons, the data were not analyzed and are a limitation in this study design. Future research should include a more comprehensive and standardized assessment of substances known to affect performance.

Although the Apple Watch has been validated in other studies [7,18,19,53], there were a number of possible limitations. Although Apple does not promote a specific sleep analysis feature, heart rate data have been known to correlate with sleep patterns, and recent research has begun to validate the use of raw Apple Watch heart rate data to derive sleep duration [54]. One limitation was the range of heart rates that the Apple Watch is technically able to collect, which consists of a minimum of 30 BPM and a maximum of 210 BPM, as highlighted in Table 3 [37]. For instance, an individual could register a heart rate of 30 BPM, such as a physiological pause after a premature beat. Similarly, a high heart rate recorded at 210 BPM would not necessarily mean that the heart rate was sustained at this level but that it was sustained enough to register at the upper limit. The frequency of low or high heart rate events depends on the frequency of sampling at a higher rate in workout mode (every 5 seconds) versus regular mode (every 5 min). Therefore, the mean heart rates are more representative of our cohort’s data. The fact that we reported within our range values both very low and very high step counts and energy expenditures is a result of a number of factors. If a trainee was removed from training early, spent the day in a classroom, or did not enter the workout mode, this would account for low step count and energy expenditure. For these reasons, we feel that the mean values are more reflective of the data. However, to control for these limitations, an additional and different wearable sensor with a wider range of data collection capabilities can be used in future studies in conjunction with the Apple Watch to further compare data.

Owing to the nature of military training, another limitation is that the number of subjects per group is highly unequal. To address this, we combined data from the 3 cohorts across groups to create larger sample sizes. In the future, we will require additional cohorts to increase the sample size and validate the results.

Another possible limitation of our study is the use of self-report surveys, which can be prone to bias or misreporting. However, other studies in military populations have used the same or similar personality and behavioral self-report surveys [8,11-13,15,23,25,26]. Although we agree that self-assessment is not generalizable, our ability to obtain daily individual measures of mental and behavioral status, we feel, was a great strength in this study. Furthermore, we recognize that all observational research has the potential for bias to some degree, but we feel as though this research is crucial in contributing to human performance knowledge by not disrupting the natural training environment of these elite warfighters.

Conclusions

This study demonstrates the feasibility and accuracy of actionable data that can be collected using a natural, continuous, and holistic model of data collection using a modern digital platform, custom-made software, and body-worn sensors. This model of innovation has the potential for rapid discovery in military training environments that may lead to better training and selection of military personnel and translate into other elite training environments. In addition, the data paint a novel picture of the mind and body processes in determining performance outcomes—findings that can generalize into any high pressure, competitive domain.

Acknowledgments

The authors are deeply indebted to the leadership of the Reconnaissance Training Company at the School of Infantry-West (SOI-W) at Camp Pendleton for their partnership, ideas, and numerous hours spent teaching the authors their objectives and needs. These crucial interactions enabled the authors to design a study targeted at meeting real objectives efficiently. The authors are also grateful to the leadership of the SOI-W for supporting this line of research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Daily survey questions.
[DOCX File, 13 KB - mhealth_v8i6e14116_appl1.docx ]

Multimedia Appendix 2

Rates of successful course completion and failure to complete by category for all 3 classes. DOR: drop on request.
References


32. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag J Sci 2009 Apr 21;50(302):157-175. [doi: 10.1080/178640009463897]


Abbreviations

- **BFI**: Big Five Inventory
- **BPM**: beats per minute
- **BRC**: Basic Reconnaissance Course
- **BRPC**: Basic Reconnaissance Primer Course
- **BUD/S**: Basic Underwater Demolition/SEAL
- **DOR**: drop on request
- **ECG**: electrocardiogram
- **ERS**: Ego Resilience Scale
- **FFMQ**: Five Facet Mindfulness Questionnaire
- **PANAS**: Positive and Negative Affect Scale
- **SEAL**: Sea, Air, and Land
- **SOI-W**: School of Infantry-West
- **SWLS**: Satisfaction with Life Scale

©Leslie Saxon, Brooks DiPaula, Glenn R Fox, Rebecca Ebert, Josiah Duhaime, Luciano Nocera, Luan Tran, Mona Sobhani. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 15.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
A Nonproprietary Movement Analysis System (MoJoXlab) Based on Wearable Inertial Measurement Units Applicable to Healthy Participants and Those With Anterior Cruciate Ligament Reconstruction Across a Range of Complex Tasks: Validation Study

Riasat Islam¹, BSc, MSc; Mohamed Bennasar¹, PhD; Kevin Nicholas²,³, BSc; Kate Button²,³, PhD; Simon Holland¹, PhD; Paul Mulholland⁵, PhD; Blaine Price¹, MSc; Mohammad Al-Amri²,⁴, PhD

¹School of Computing and Communications, The Open University, Milton Keynes, United Kingdom
²Biomechanics and Bioengineering Research Centre Versus Arthritis, Cardiff University, Cardiff, United Kingdom
³Cardiff and Vale University Health Board, Cardiff, United Kingdom
⁴School of Healthcare Sciences, College of Biomedical and Life Sciences, Cardiff University, Cardiff, United Kingdom
⁵Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom

Corresponding Author:
Mohammad Al-Amri, PhD
School of Healthcare Sciences
College of Biomedical and Life Sciences
Cardiff University
Room 13.17, 13th Floor
Eastgate House, 35-43 Newport Road
Cardiff, CF24 0AB
United Kingdom
Phone: 44 2920687115
Email: al-amrim@cardiff.ac.uk

Abstract

Background: Movement analysis in a clinical setting is frequently restricted to observational methods to inform clinical decision making, which has limited accuracy. Fixed-site, optical, expensive movement analysis laboratories provide gold standard kinematic measurements; however, they are rarely accessed for routine clinical use. Wearable inertial measurement units (IMUs) have been demonstrated as comparable, inexpensive, and portable movement analysis toolkits. MoJoXlab has therefore been developed to work with generic wearable IMUs. However, before using MoJoXlab in clinical practice, there is a need to establish its validity in participants with and without knee conditions across a range of tasks with varying complexity.

Objective: This paper aimed to present the validation of MoJoXlab software for using generic wearable IMUs for calculating hip, knee, and ankle joint angle measurements in the sagittal, frontal, and transverse planes for walking, squatting, and jumping in healthy participants and those with anterior cruciate ligament (ACL) reconstruction.

Methods: Movement data were collected from 27 healthy participants and 20 participants with ACL reconstruction. In each case, the participants wore seven MTw2 IMUs (Xsens Technologies) to monitor their movement in walking, jumping, and squatting tasks. The hip, knee, and ankle joint angles were calculated in the sagittal, frontal, and transverse planes using two different software packages: Xsens’ validated proprietary MVN Analyze and MoJoXlab. The results were validated by comparing the generated waveforms, cross-correlation (CC), and normalized root mean square error (NRMSE) values.

Results: Across all joints and activities, for data of both healthy and ACL reconstruction participants, the CC and NRMSE values for the sagittal plane are 0.99 (SD 0.01) and 0.042 (SD 0.025); 0.88 (SD 0.048) and 0.18 (SD 0.078) for the frontal plane; and 0.85 (SD 0.027) and 0.23 (SD 0.065) for the transverse plane (hip and knee joints only). On comparing the results from the two different software systems, the sagittal plane was very highly correlated, with frontal and transverse planes showing strong correlation.
Conclusions: This study demonstrates that nonproprietary software such as MoJoXlab can accurately calculate joint angles for movement analysis applications comparable with proprietary software for walking, squatting, and jumping in healthy individuals and those following ACL reconstruction. MoJoXlab can be used with generic wearable IMUs that can provide clinicians accurate objective data when assessing patients’ movement, even when changes are too small to be observed visually. The availability of easy-to-set up, nonproprietary software for calibration, data collection, and joint angle calculation has the potential to increase the adoption of wearable IMU sensors in clinical practice, as well as in free living conditions, and may provide wider access to accurate, objective assessment of patients’ progress over time.

DOI: 10.2196/17872

KEYWORDS

gait; algorithms; motion trackers; lower extremity; wearable electronic devices; exercise therapy; digital physiotherapy; kinematics; wearables; range of motion; anterior cruciate ligament

Introduction

Within biomechanics, sports science, and physiotherapy, assessment of movement patterns for activities such as walking is vital to their practice to inform decision making around performance, recovery, and risk of reinjury [1]. In clinical settings, physiotherapy practice at present relies extensively on the visual assessment of movement quality and on the use of associated subjective clinical scales [2]. Both play important roles in decision making for treatment selection. However, there can be considerable variation in the quality of assessments depending on the experience of physiotherapists, and interrater agreement is not always as strong as expected [2]. An objective and more accurate assessment during physiotherapy sessions has the potential to facilitate more accurate diagnoses and more consistent treatment selection [3]. It also has the potential to provide objective feedback to surgeons to demonstrate the actual postoperative effect of different decisions made during surgery [4]. This is very important for patients with anterior cruciate ligament (ACL) reconstruction. ACL rupture is a common sporting injury to the knee that frequently results in surgery to reconstruct the ligament [5,6]. This is followed by a lengthy period of rehabilitation, and the ability of these individuals to return to sports varies and has been reported to be as low as 65% returning to their preinjury level of activity [7]. The reasons for this are multifactorial, but an important factor is that people with ACL reconstruction are known to move with biomechanical compensation strategies despite rehabilitation [8-10]. This can put them at risk of reinjury and future osteoarthritis, so it is important that clinicians have tools available to them in the clinical setting to assess the biomechanics during tasks that mimic sporting maneuvers [10-12].

3D motion capture camera–based systems can provide a gold standard assessment of body movement; however, such laboratories are expensive, time consuming, labor intensive, and effectively nonportable [13]. Data analysis is similarly resource intensive, time consuming, and requires specially trained personnel [14]. These limit 3D motion capture camera–based systems to research settings and are scarce in clinical practice [1]. However, wearable inertial measurement units (IMUs) can offer a similar objective assessment of body movement and are relatively much less expensive, easier to setup, mobile, and usable by clinicians with minimal training. IMU sensors consist of a triaxial accelerometer, a triaxial gyroscope, and a triaxial magnetometer. Data collected by an IMU is processed to calculate the sensor position, speed, and orientation. For certain IMUs and software, these results have been shown to be comparable to 3D motion capture camera–based systems [3]. These characteristics strongly suggest that sensors have great potential for use in clinical practice. The availability of validated and low-cost nonproprietary systems could make such systems affordable and much more widely used in clinical practice.

Existing systems pose a limitation of having complex calibration processes. Hullfish et al [2] attempted to address this issue by presenting a self-calibrated wearable sensor system for knee joint angle measurements only. Even though they have used a single, low-cost wearable inertial sensor and a simple calibration process, the system is not suitable for more complex activities such as walking, squatting, and jumping. Moreover, they have not demonstrated the use of the system in a clinical setting or included any patients. Similarly, Nazarahari et al [15] proposed a calibration method using multiple wearable IMUs to reduce measurement errors due to calibration for gait kinematics. The proposed calibration method [15] is simpler than some of the existing methods; however, the calibration requires specific movements such as hip abduction and adduction to a predefined degree, which might be a challenge for people with knee conditions.

To address the limitations mentioned earlier, we have developed nonproprietary software, MoJoXlab [16], through an academic–clinical research collaboration. MoJoXlab [16] has been developed to provide a more practical system for clinical movement analysis. The software can be used with any generic wearable IMUs that produces orientation angles in quaternions. It employs a simple protocol for data collection and calibration to facilitate the use of wearable IMU sensors for clinical movement analysis as the users deem fit and can also be used for diagnosis and prognosis in clinical settings. MoJoXlab [16] implements an IMU-to-body calibration method [17-19]. Although previous studies [17] have explored this method during simple activities, such as walking in healthy participants, it is yet to be explored during complex activities such as squatting or jumping, which are of interest to clinicians rehabilitating people back to sports. Within the movement analysis domain, jumping is considered to be a complex activity because of its dynamic nature, such that even conventional gait measurement equipment finds it difficult to measure accurately.
The data obtained from wearable IMU sensors deviate significantly from 3D motion capture camera–based data, owing to the large impact on ground contact. The proprietary software MVN Analyze (Xsens Technologies) solves this issue to a certain extent [20]. Currently, it is the only available validated software system and as a result has been used in this research as the gold standard. However, it is limited to only Xsens’ proprietary IMU hardware. Therefore, there is a need to develop a software system that can be used with any suitable IMU.

The aim of this study was to compare the hip, knee, and ankle joint angles calculated in the sagittal, frontal, and transverse planes by MoJoXlab [16] against MVN Analyze (Xsens Technologies) from movement data collected using wearable IMU sensors during walking, squatting, and jumping in healthy people in a nonclinical setting and people with ACL reconstruction in a clinical setting.

### Table 1. Participant demographics.

<table>
<thead>
<tr>
<th>Participants</th>
<th>Sample, N</th>
<th>Average age (years)</th>
<th>Gender (male, female)</th>
<th>Average Height (cm)</th>
<th>Average Weight (kg)</th>
<th>Knee injury (right, left)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>27</td>
<td>35 (SD 9)</td>
<td>11, 16</td>
<td>162 (SD 34)</td>
<td>72 (SD 13)</td>
<td>0</td>
</tr>
<tr>
<td>ACL</td>
<td>20</td>
<td>29 (SD 9)</td>
<td>14, 6</td>
<td>177 (SD 11)</td>
<td>84 (SD 18)</td>
<td>8, 12</td>
</tr>
</tbody>
</table>

**Experimental Protocol**

Each participant underwent at least one movement analysis session, with all healthy participants returning for another session within about a week later (mean 4, SD 3 days). On each day, the measurements were repeated twice, once with a biomechanics expert putting on the sensors and performing data collection, and in the other instance, a physiotherapist performed the same. Thus, a total of 4 sessions of data were recorded for each healthy participant (day 1—experiment 1 [performed by a biomechanics expert], day 1—experiment 2 [performed by a physiotherapist]; day 2—experiment 1 [performed by a biomechanics expert]; day 2—experiment 2 [performed by a physiotherapist]). The interrater reliability was acceptable across all planes for walking and squatting joint angles and for jumping it ranged from poor to excellent [3].

Data collection for all healthy participants was carried out in the Research Center for Clinical Kinaesiology at Cardiff University. In the case of participants with ACL reconstruction, data collection took place in the clinic and was conducted by a physiotherapist. For all the participants (both healthy and people with ACL reconstruction), during the first session, anthropometric measurements were taken by a physiotherapist from the right lower limb while the participant maintained a standing posture. A total of 7 MTw2 trackers (Xsens Technologies, Enschede, Netherlands) were then placed in accordance with Xsens’ instructions [21]. MTw2 trackers were secured using elasticated Velcro straps on each upper thigh (centrally and halfway between the greater trochanter and lateral epicondyle of the knee), each lower leg (proximal medial surface of the tibia), the dorsum of each foot, and one centrally over the sacrum. Each lower limb tracker was placed between the two outermost layers of the strap and attached to the Velcro of the inner layer to secure its position and minimize any movement. The sacral tracker was placed directly over the sacrum, with the upper border of the sensor aligned centrally between the two posterior superior iliac spines. The sacral sensor was held in position with medical grade double-sided adhesive tape.

Wherever possible, all the participants (both ACL and healthy) performed 8 repetitions of each of the following 3 activities: overground walking, squatting, and vertical jumping. Some of the ACL reconstruction participants were exempted from activities that they found difficult, for example, only 7 participants performed the vertical jump. Before performing each activity, the participant was provided with a demonstration by the physiotherapist and could ask any questions. The order of the activities was randomized across participants, but consistent within participants. Each walking trial consisted of a walk in a straight line across the laboratory or clinic (approximately 8 m) at the participants’ natural pace. For healthy participants, the walking trial was repeated 5 times, and for participants with ACL reconstruction, the walking was repeated 2 times. A walking trial consisted of 8 gait cycles of walking, and similarly, one such jumping or squatting trial consisted of 8 jumps and squats. The squat depth and jump height were not measured.

**MoJoXlab**

MoJoXlab [16] is a MATLAB-based (version 2018b; The MathWorks Inc) custom motion capture analysis software toolkit whose aim is to produce freely available motion capture analysis software to be used by anyone interested in generating lower limb joint kinematic waveforms using any suitable IMUs [16]. MoJoXlab [16] was used in this study to generate joint angles for different functional tasks such as walking, squatting, and jumping. The joint angles are then validated against the commercially available MVN Analyze (Xsens Technologies). The joint kinematics for MoJoXlab [16] are based on a joint coordinate system, as proposed by Grood and Suntay [18].

**Methods**

**Research Participants and Setting**

Ethical approval for this study was obtained from Wales Research Ethics Committee 3 (10/MRE09/28). Written informed consent was obtained before participation. A sample of healthy participants (n=27) was recruited using the following criteria: age between 18 and 60 years; healthy with no known neurological, cardiovascular, or musculoskeletal conditions. Additionally, 20 participants who underwent ACL reconstruction were recruited from physiotherapy and orthopedic knee clinics in one University Health Board using the following criteria: age between 18 and 60 years, had ACL surgery in at least one of their knees within 6 to 12 months. Participant demographics are presented in Table 1.
MoJoXlab [16] takes sensor orientation data in quaternions as input and output joint angles in degrees. The joint angles generated by this method were then compared with the joint angles generated by the proprietary MVN Analyze software (Xsens Technologies). The algorithm considers a static calibration step, where sensor data are captured for calibration purposes while the participant maintains a standing pose [17]. This calibration step was then used to calculate the joint angles for the dynamic phase of the motion.

**Calibration and Data Collection**

Kinematic data were collected using the MTw2 trackers at 60 Hz, and all the trackers were connected to the computer using Wi-Fi technology. The data were recorded on a computer using the Xsens MVN Analyze system (Xsens Technologies). Before beginning the tasks, the participant was asked to stand in a static N-pose, as per the instructions in the MVN Analyze user manual [21]. This was maintained for approximately 30 seconds. At the start of this period of quiet stance, the MTw2 trackers were calibrated using the MVN Analyse software (Xsens Technologies). During this process, the software establishes the relationship between the body segment and tracker orientations [18,22]. The calibration data saved within MVN Analyze (Xsens Technologies) are proprietary and cannot be extracted. Consequently, additional static calibration data were collected by asking the participant to maintain a standardized standing posture, to be used as a static calibration dataset within our custom MoJoXlab software [16,17]. This allows raw data collected from all MTw2 trackers to be projected to one global coordinate system. Then, the data for each activity were collected using MVN Analyze software (Xsens Technologies). The purpose of using the MVN Analyze software (Xsens Technologies) was two-fold: to capture the data streamed from the trackers and to later calculate the joint angles as per MVN’s proprietary algorithm to compare the joint angles with our custom MoJoXlab software [16]. However, MoJoXlab [16] uses the raw data from the same trackers to calculate joint angles and is independent of the MVN Analyze software (Xsens Technologies).

**Data Processing**

As mentioned earlier, data obtained from the trackers were saved using the MVN Analyze software (Xsens Technologies). Hip, knee, and ankle joint angle calculations were also performed using the same proprietary software. All the data generated by MVN Analyze (Xsens Technologies) were exported in mvnx file formats (MVN Analyze’s open XML data format). These files were later imported to the MATLAB software (version 2018b; The MathWorks Inc), and MoJoXlab [16] was used to extract the raw sensor data from the mvnx files to calculate another set of hip, knee, and ankle joint angles.

The joint angles were generated for each activity (walk, jump, and squat), each joint (hip, knee, and ankle), each plane of movement (sagittal, frontal, and transverse), and each side of the body (left and right). However, our custom algorithm within MoJoXlab [16] could only generate angles in the sagittal and frontal planes of the ankle joint. Positive joint angles indicate flexion, abduction, and internal rotation in the sagittal, frontal, and transverse planes, respectively.

**Data Analysis and Validation**

Joint angles obtained by MVN Analyze (Xsens Technologies) and MoJoXlab [16] were compared and analyzed using separate custom scripts written in MATLAB (version 2018b; The MathWorks Inc). The workflow for data processing, analysis, validation, and visualization is outlined in Figure 1.

Custom MATLAB (version 2018b; The MathWorks Inc) scripts were used as follows:

1. Extract joint angles calculated by MVN Analyze (Xsens Technologies) from mvnx files and then saved as the MAT file (MATLAB’s data format).
2. Raw sensor data in quaternions from mvnx files were extracted, and MoJoXlab [16] was used to calculate another set of joint angles distinct from those calculated by MVN Analyze (Xsens Technologies). The joint angles were then saved to the MAT file.
3. Visualize joint angle waveforms and calculate cross-correlation (CC) and root mean square error (RMSE) values between the waveforms and plot their graphs.

During the data collection phase for each healthy participant, 4 trials were collected, so for 27 healthy participants, a total of 108 trials were collected. Of these, 13 were excluded from the analysis because there were some data missing for each of them. For healthy participants, 95 data trials were used for the analysis.

Similarly, for 20 ACL reconstruction participants, only one of them returned for a repeat session. All of them performed walking and squatting activities, but only 7 performed the jumping activity. Data for a total of 21 walk and squat trials were collected, and 8 jump trials were collected (1 participant performed the jump repeat session). Out of the 21 walk and squat trials, 2 were excluded as some data were missing. Thus, a total of 19 trials were used in the data analysis for the walk and squat activities, and 8 jump trials were used in the data analysis.

Waveforms of joint angles generated by MVN Analyze (Xsens Technologies) and MoJoXlab [16] were compared inside the MATLAB (version 2018b; The MathWorks Inc) environment using custom scripts, as explained earlier. In a previous study, the joint angles obtained from MVN Analyze (Xsens Technologies) were validated against joint angles obtained from the gold standard Vicon system (Vicon Motion Systems Ltd) [3]. Thus, in this study, the joint angles generated from MVN Analyze (Xsens Technologies) can be used as reference values to compare the joint angles generated by MoJoXlab [16].
Cross-Correlation

CC is a similarity metric used in signal processing to assess the similarity between two signals [23-25]. The resultant values are obtained as vectors. By using the `coeff` function in MATLAB (version 2018b; The MathWorks Inc), it is possible to calculate the CC coefficient between the two compared signals [26]. The metric can then be interpreted in a similar manner to the Pearson correlation coefficient, producing values between 0 and 1, with values closer to 1 indicating a higher correlation between the signals, and thus greater similarity.

CC between MoJoXlab [16] and MVN Analyze (Xsens Technologies) was calculated for each waveform to test for similarity. First, the waveforms were center normalized to have a mean of zero and corrected for polarity. The CC coefficient was calculated in the range of 0 to 1, with values closer to 1 indicating a very high correlation.
Normalized Root Mean Square Error

RMSE is an alternative method of measuring the differences between sets of values. In this study, RMSE was used to measure the error in joint angle values between MVN Analyze (Xsens Technologies) and MoJoXlab [16]. Two details are worth mentioning: first, applying RMSE naively to joint angles would produce error values in degrees. However, owing to the wide variety of joints, tasks, and participant groups, it would be difficult to compare RMSE values in a meaningful way across the dataset. To address this problem, the normalized version of the RMSE was used. This produces the values within the range of 0 to 1, where closer to 0 indicates a lower error (better agreement) between the joint angle waveforms.

Normalized root mean square error (NRMSE) was calculated to compare the joint angles between MoJoXlab [16] and MVN Analyze software (Xsens Technologies) [25,27-29]. The waveforms were standardized over the range of 0 to 1 and corrected for polarity. NRMSE was obtained within the range of 0 to 1, where values closer to 0 indicate the least difference between the waveforms.

Results

Waveforms

This section presents joint angle waveforms, generated by MVN Analyze software (Xsens Technologies) and MoJoXlab [16], across all movement planes (sagittal, frontal, and transverse), for each joint (hip, knee, and ankle) and for each task (walk, squat, and jump). Figures 2-4 show representative joint angle waveforms from the healthy participant dataset. Figures 5-7 show representative joint angle waveforms from the people with ACL reconstruction dataset.

Figure 2. Representative sagittal plane joint angle waveforms from the healthy participant data set. Waveforms for the hip (top row), knee (middle row), and ankle (bottom row) joint angles obtained from MVN Analyze (blue) and our custom software MoJoXlab (orange) for walking (left), squatting (center), and jumping (right) tasks. The y-axis represents joint angles in degrees, and the x-axis represents data samples across the entire waveform.
Figure 3. Representative frontal plane joint angle waveforms from the healthy participant data set. Waveforms for the hip (top row), knee (middle row), and ankle (bottom row) joint angles obtained from MVN Analyze (blue) and our custom software MoJoXlab (orange) for walking (left), squatting (center), and jumping (right) tasks. The y-axis represents joint angles in degrees, and the x-axis represents data samples across the entire waveform.

Figure 4. Representative transverse plane joint angle waveforms from the healthy participant data set. Waveforms for the hip (top row) and knee (bottom row) joint angles obtained from MVN Analyze (blue) and our custom software MoJoXlab (orange) for walking (left), squatting (center), and jumping (right) tasks. The y-axis represents joint angles in degrees, and the x-axis represents data samples across the entire waveform.
Figure 5. Representative sagittal plane joint angle waveforms selected from the anterior cruciate ligament reconstruction participant data set. Waveforms for the hip (top row), knee (middle row), and ankle (bottom row) joint angles obtained from MVN Analyze (blue) and our custom software MoJoXlab (orange) for walking (left), squatting (center), and jumping (right) tasks. The y-axis represents joint angles in degrees, and the x-axis represents data samples across the entire waveform.

Figure 6. Representative frontal plane joint angle waveforms selected from the anterior cruciate ligament reconstruction participant data set. Waveforms for the hip (top row), knee (middle row), and ankle (bottom row) joint angles obtained from MVN Analyze (blue), and our custom software MoJoXlab (orange) for walking (left), squatting (center), and jumping (right) tasks. The y-axis represents joint angles in degrees, and the x-axis represents data samples across the entire waveform.
Validation Results

This section presents the validation results for the joint angle waveforms using CC and NRMSE. The MoJoXlab [16] joint angle waveforms were compared with waveforms generated by MVN Analyze (Xsens Technologies). CC and NRMSE values were calculated for each task, each joint, and each plane. The results are presented later in parts. First, the CC values are shown for both healthy and ACL reconstruction participants. Afterward, NRMSE values are presented for healthy and ACL reconstruction participants.

The mean CC values across all participants (healthy and ACL reconstruction participants) were very high (CC>0.95) for the sagittal plane across all the joints and tasks. For healthy participants in the frontal plane across all tasks, CC>0.83, and for ACL reconstruction participants for the frontal plane across all tasks, CC>0.78. Similarly, for the transverse plane, for healthy participants across all tasks, CC>0.83, and for ACL reconstruction participants across all tasks, CC>0.84.

The NRMSE for the sagittal plane was relatively low compared with other planes for all participants (healthy and ACL across all tasks: NRMSE<0.1). For healthy participants in the frontal plane across all tasks, NRMSE<0.17, and for ACL reconstruction participants for the frontal plane across all tasks, NRMSE<0.35. Similarly, for the transverse plane, for healthy participants across all tasks, NRMSE<0.22, and for ACL reconstruction participants across all tasks, NRMSE<0.39.

Discussion

Principal Findings

This paper has demonstrated that MoJoXlab [16], our in-house developed software, can be used to calculate joint angles for movement analysis with generic wearable IMUs that report data in quaternions. MoJoXlab [16] has a simple calibration procedure, making the data collection process smooth. This makes MoJoXlab [16] potentially easier to use in clinical settings, and this paper has established its validity and demonstrated that MoJoXlab [16] can be used in a clinical setting by a clinician, across a variety of complex tasks such as walking, squatting, and jumping, and across a variety of participants, both healthy and ACL reconstruction participants. Complex tasks such as jumping are very challenging to analyze accurately with wearable IMU sensors because of the large ground impact force. MoJoXlab [16] can accurately calculate joint angles for such complex tasks and thus can be potentially extended to calculate other complex tasks and exercises as well. MoJoXlab [16] has been validated against proprietary MVN Analyze software (Xsens Technologies), which was previously validated against the VICON-based optical motion capture.
system (Vicon Motion Systems Ltd), considered to be clinically gold standard [3].

Al-Amri et al [3] concluded that joint angle waveforms obtained from MVN Analyze (Xsens Technologies) showed excellent similarity with sagittal plane waveforms obtained by the VICON system (Vicon Motion Systems Ltd) and acceptable similarity for frontal and transverse planes across all three tasks. MVN Analyze (Xsens Technologies) and VICON systems (Vicon Motion Systems Ltd) were compared using the coefficient of multiple correlation (CMC) and $R^2$ values for the linear fit method. The CMC was found to be greater than 0.9 for all three joints in the sagittal plane across all tasks. Similarly, for the sagittal planes, the $R^2$ value was greater than 0.8 for all the joints across all the tasks, and similarly $R^2$ values showed fair-to-good similarity for transverse and frontal planes across all joints during squatting and jumping and knee joint during walking. Thus, by the transitive property, we claim that MoJoXlab [16] can generate joint angles comparable with optical gold standard motion capture systems.

In the following sections, we discuss the validation results between MoJoXlab [16] and MVN Analyze (Xsens Technologies) for each of the planes in two ways: by comparing the joint angle waveforms across CC and by computing the NRMSE. We also discuss differences in healthy participants versus ACL reconstruction participants across activities.

**Cross-Correlation**

For all joints, across all tasks and participants (both ACL reconstruction participants and healthy participants), the sagittal plane shows a very high correlation, with mean CC above 0.95. This indicates that MoJoXlab [16] generates sagittal plane joint angle waveforms that are highly similar to those of MVN Analyze (Xsens Technologies). The sagittal plane reflects joint angles for flexion and extension of the joints, which are most commonly referred by clinicians [30], to assess recovery and potential risk factors for injury to the ACL. Reduced range of motion in this plane is often associated with incomplete recovery and poor neuromuscular control. For example, reduced knee flexion during landing from a jump is associated with higher peak moments at the knee joint [31].

Similarly, the frontal plane is also useful for clinicians who are interested in abduction and adduction of the joints, as this is considered a risk factor for reinjury, poor neuromuscular control, and incomplete recovery [30,32]. In the case of frontal planes, CCs are also high for all joints: with values across all tasks and participant groups for ankle joints being greater than 0.84, for hip joints being greater than 0.78, and for knee joints being greater than 0.83.

In the case of the transverse plane, MoJoXlab [16] can calculate joint angles for the hip and knee joints only. In this plane, CC values across all tasks and participant groups for the hip are greater than 0.83 and for knee joints, greater than 0.83.

Overall, by observing the representative waveforms (Figures 2-7) and the high CC values (Figure 8), it is evident that MoJoXlab [16] software can produce joint angles comparable with the commercial MVN Analyze software (Xsens Technologies).

Previous studies comparing software to calculate joint angles using wearable IMUs are limited. Hullfish et al [2] investigated knee joint angles in the sagittal plane only for seven healthy participants. They compared their IMUs with an optical motion capture system. Their CC values were within the range of 0.84 to 0.99. In comparison with these values, we obtained a mean CC range of greater than 0.95 for the sagittal plane across all participant groups, activities, and all joints. For other planes, the CC is generally greater than 0.83 except for the frontal plane for ACL reconstruction participants, where the values are greater than 0.78.

These results are comparable with those of previous studies and further extend previous work in healthy participants, which reported high agreement between joint angle waveforms in the sagittal plane for systems using IMUs and an optical motion capture system [33,34]. Other studies have compared data obtained from Xsens IMUs for walking [35,36], squatting [14,37], and jumping [38,39]. However, our results extend previous work by including more challenging dynamic tasks such as squatting and jumping. We also evaluated the validity of software in people with ACL reconstruction in addition to healthy people. The results confirm that MoJoXlab software [16] can be used to assess tasks such as squatting and jumping in healthy individuals and individuals following ACL reconstruction within a clinical setting.
Figure 8. Mean cross-correlation values between MVN Analyze and MoJoXlab for healthy participants (black) and anterior cruciate ligament reconstruction participants (white), presented in a graphical format for visualization and overall comparison purposes. Values close to 1 indicate a very high correlation (circle—values for the walk task, square—values for jump task, and diamond—values for squat task; planes: Fnt—frontal, Sag—sagittal, and Tran—transverse).

Normalized Root Mean Square Error

CC as a measure of similarity is blind to both constant vertical offsets and differences in amplitude. It should be noted that blindness to constant vertical differences are not of material interest for the purposes of the study; however, differences in amplitude are of considerable importance because they represent entirely different joint angle ranges. For this reason, it is valuable to use a complementary measure of similarity, which is highly sensitive to differences in amplitude. In particular, the RMSE corresponds to a single number representing the Pythagorean distance in a high-dimensional space between the two waveforms and is highly sensitive to differences in amplitude, frequency, and offset. To allow meaningful comparison of RMSE values between different activities and joints, we have given the results as NRMSE, where the RMSE is divided in each case by the range (Figure 9).

As noted previously, CC values for sagittal planes showed very high agreement between the waveforms generated by the two systems. Similarly, the NRMSE values obtained for sagittal planes also showed a very low error (NRMSE<0.1) across all tasks, joint angles, and participant groups. The low NRMSE values in conjunction with very high CC values suggest that MoJoXlab [16] can generate joint angle waveforms in the sagittal plane that are highly comparable with commercially available MVN Analyze software (Xsens Technologies).

In the case of frontal planes, the healthy participant joint angles showed lower error values (NRMSE<0.17) than the ACL reconstruction participants group (NRMSE<0.35). Similarly, in the transverse plane, the healthy participant joint angles showed lower error values (NRMSE<0.22) than the ACL reconstruction participants group (NRMSE<0.39). Thus, the NRMSE values for the ACL reconstruction participants group for both the frontal and transverse planes were higher than their respective healthy participant group values. The error values for joint angles for the frontal and transverse planes for healthy participants are within the reasonably accepted range of 0.2. The high CC and low NRMSE values for all healthy participants across all tasks and joints suggest excellent agreement between MoJoXlab [16] and MVN Analyze (Xsens Technologies).

In summary, in the case of the ACL reconstruction participant group, values for the sagittal plane show high CC and low NRMSE values, suggesting excellent agreement. For ACL reconstruction participants, in the transverse and frontal planes, the CC values are high, thus confirming agreement on waveform pattern similarities between MoJoXlab [16] and MVN Analyze (Xsens Technologies). However, the underlying reasons behind the slightly higher range of NRMSE error values for the ACL reconstruction participants group for frontal planes and transverse planes requires further investigation.
Understanding the Differences in Waveforms

The differences in joint angle waveforms for the same task and joint noted above may be due to several possible contributing factors. One of the significant contributing factors is the static calibration step described in the Methods section. The calibration step carried out by the proprietary MVN system produces no externally inspectable data that can be used in this study. The calibration values captured by MVN during the calibration step were saved internally within the software. The values are not accessible to the user either on the software interface or when all the data are exported as *.mvnx files. Thus, MoJoXlab [16] does not have access to MVN calibration values. A separate set of values were captured for MoJoXlab [16] as its calibration step, while the participant maintained the same standing posture. In principle, the calibration values should be similar to the data being captured for the same standing posture. However, it is reasonably possible that minuscule movements can vary the calibration values, even more so for ACL reconstruction participants than for healthy participants. It is likely that ACL reconstruction participants might find it difficult to maintain the same standing posture while the calibration steps are carried out. This might be a contributing factor to the difference in waveforms between the two software systems and also for the slightly larger NRMSE values or the ACL reconstruction participants in comparison with healthy participants. In contrast, the healthy participants might have held similar static postures for the static calibration step, resulting in similar calibration values feeding to the two software systems. As a result, the waveforms are more in agreement for this participant group.

Another potential contributing factor is the different sites for data collection. For the healthy participant group, both the static calibration steps and the functional tasks were measured in the laboratory. However, in the ACL reconstruction participants group, the calibration step was undertaken in the consultation room within the clinic, and some of the activities took place outside of the consultation room in the corridor or in a different room. The different sites for data collection for some of the tasks can account for the difference in the waveforms as a number of external factors can contribute to the difference in waveforms between the two software systems. One such external factor is the presence of equipment in the clinic that causes magnetic interference. The physiotherapist conducting the data collection in the clinic noted that there was external magnetic interference affecting the sensor data. In principle, as the two software packages use the same raw sensor data, it can be understood that magnetic interference should not affect the outcome of the joint angle waveforms. However, there can be a difference in waveforms because MoJoXlab [16] and MVN Analyze (Xsens Technologies) handle magnetic interference in different ways. Currently, MoJoXlab [16] does not have any

Figure 9. Normalized root mean square error values between MVN Analyze and MoJoXlab for healthy participants (black) and ACL reconstruction participants (white), presented in a graphical format for visualization and overall comparison purposes. Values close to 0 indicate a very low error (circle—values for the walk task, square—values for jump task, and diamond—values for squat task; planes: Frnt—Frontal, Sag—Sagittal, and Tran—Transverse).
special software or algorithm that handles magnetic interference from the environment, and this is a limitation of the current version of MoJoXlab [16]. However, Xsens claims that they have special software in MVN Analyze to handle magnetic interference from the environment, even though such claims have not yet been validated [40]. To use MoJoXlab [16] with data collected in clinical settings, people should be careful to determine whether magnetic interference severely affects the data. It is possible that discrepancies could occur when collecting data over time due to external magnetic interference [41].

In clinical settings, 3D motion capture camera–based systems are generally used for movement analysis, which tracks markers attached to the body over a certain field of space. As a result, it is possible to detect the movement of the body frame and body segments across time and space. However, motion tracking using IMUs uses an inherently different principle, where the relative angular motion of each IMU sensor is combined using sensor fusion algorithms to calculate the joint angles for the hip, knee, and ankle joints. Thus, one of the limitations of using IMUs for clinical movement analysis is that joint angles are considered separately for each joint; thus, phenomena such as shifting of the knee cannot be detected by simply considering joint angles [42].

One of the major limitations of this study is that, while the number of trials available for analysis in the healthy participant group was quite large at 96, the number was relatively small for ACL reconstruction participants, with walking and squatting tasks having 19 available trials, whereas the jumping task had only eight available trials. This disparity in the number of trials available for data analysis between healthy and ACL reconstruction participants can also be one of the contributing factors to the difference in results. Further work is required to collect more data from people with ACL reconstruction for a better comparison between healthy and ACL reconstruction participants’ data.

Further Work
MoJoXlab [16] is currently under development in collaboration with Cardiff University and the Open University. This paper presented only the validation results between the waveforms generated by MoJoXlab [16] and the proprietary MVN software. Further work is required to validate the various gait parameters calculated by MoJoXlab [16] and to enable MoJoXlab [16] to better handle external factors that can affect the data, such as magnetic interference. Although MoJoXlab [16] can work with any sensor that reports quaternions, in this particular study, the IMU data were collected using Xsens’ wearable IMU sensors. In the future, we would like to test how different wearable IMU sensors can be used with MoJoXlab [16].

Conclusions
This study has shown that a variety of clinically relevant functional tasks such as walking, squatting, and jumping can be measured using wearable IMUs in both laboratory and clinical settings, by clinicians (in this case physiotherapists) using nonproprietary software. We have developed and validated this nonproprietary software against software that has been shown to be as accurate as an optical motion capture system. Validation results suggest that MoJoXlab [16] can calculate joint angles comparable with proprietary MVN Analyze software (Xsens Technologies) across people with ACL reconstruction and healthy people, for tasks such as walking and more complex tasks such as squatting and jumping. Thus, MoJoXlab [16] has the potential to provide clinicians with accurate movement analysis of their patients across multiple joints and planes of motion and may be able to provide an analysis of other complex tasks such as lunging and jumping. These reflect advanced rehabilitation and sporting maneuvers that individuals following ACL reconstruction (and other injuries) need to be able to return to. It can potentially enable clinicians to benefit from using generic wearable IMUs in their practice to capture movement data of their clients and objectively track changes over time. Increasing the adoption of such software and sensors in clinical practice has the potential for better decision making around exercise prescription, monitoring patient progress over time, tailoring advice and feedback, and improving the rehabilitation process.

Acknowledgments
The authors would like to thank the volunteers for their time, enthusiasm, and feedback. KN is funded by Versus Arthritis (grant no. 18461). RI is funded by the Goldcrest Charitable Trust. MB is funded by the Engineering and Physical Sciences Research Council (grant EP/P01013X/1). BP is part funded by the Engineering and Physical Sciences Research Council (grants EP/R033862/1, EP/R013144/1, and EP/P01013X/1).

Conflicts of Interest
The MTw2 hardware and MVN Analyze software were provided on a temporary basis at no cost by Xsens Technologies (BV; the Netherlands). The authors received no financial contributions from Xsens Technologies. Training on how to use the MTw2 hardware and MVN Analyze software was provided by Xsens Technologies. Xsens Technologies had no input on data interpretation, data analysis, or manuscript writing. None of the authors have any financial interests in Xsens Technologies.

References


Abbreviations

ACL: anterior cruciate ligament
CC: cross-correlation
CMC: coefficient of multiple correlation
IMU: inertial measurement unit
NRMSE: normalized root mean square error
RMSE: root mean square error
Islam et al

Edited by G Eysenbach; submitted 17.01.20; peer-reviewed by R Richards, I Ardigò, R Ferrer-Peña, F Dittrich; comments to author 07.02.20; revised version received 30.03.20; accepted 30.03.20; published 16.06.20.

Please cite as:
Islam R, Bennasar M, Nicholas K, Button K, Holland S, Mulholland P, Price B, Al-Amri M
A Nonproprietary Movement Analysis System (MoJoXlab) Based on Wearable Inertial Measurement Units Applicable to Healthy Participants and Those With Anterior Cruciate Ligament Reconstruction Across a Range of Complex Tasks: Validation Study
JMIR Mhealth Uhealth 2020;8(6):e17872
URL: http://mhealth.jmir.org/2020/6/e17872/
doi:10.2196/17872
PMID:32543446

©Riasat Islam, Mohamed Bennasar, Kevin Nicholas, Kate Button, Simon Holland, Paul Mulholland, Blaine Price, Mohammad Al-Amri. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 16.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
The Use of a Smartphone App and an Activity Tracker to Promote Physical Activity in the Management of Chronic Obstructive Pulmonary Disease: Randomized Controlled Feasibility Study

Claire L Bentley¹, BSc, MPH; Lauren Powell², BSc, MSc; Stephen Potter¹, BSc, MSc, PhD; Jack Parker³, BSc, PhD; Gail A Mountain¹, DipCOT, MPhil, PhD; Yvonne Kiera Bartlett⁴, BSc, MSc, PhD; Jochen Farwer⁵, PhD; Cath O’Connor⁶, BSc, MSc; Jennifer Burns¹, BSc, PG Dip; Rachel L Cresswell¹, BSc, MSc; Heather D Dunn¹, BSc, MSc; Mark S Hawley¹, BSc, PhD

¹School of Health and Related Research, The University of Sheffield, Sheffield, United Kingdom
²School of Education, The University of Sheffield, Sheffield, United Kingdom
³School of Human Sciences, The University of Derby, Derby, United Kingdom
⁴Manchester Centre for Health Psychology, The University of Manchester, Manchester, United Kingdom
⁵Library, The University of Manchester, Manchester, United Kingdom
⁶Sheffield Teaching Hospitals NHS Foundation Trust, Integrated Community Care Team, Sheffield, United Kingdom

Corresponding Author:
Mark S Hawley, BSc, PhD
School of Health and Related Research
The University of Sheffield
Regent Court
30 Regent Street
Sheffield, S1 4DA
United Kingdom
Phone: 44 114 222 1726
Email: mark.hawley@sheffield.ac.uk

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is highly prevalent and significantly affects the daily functioning of patients. Self-management strategies, including increasing physical activity, can help people with COPD have better health and a better quality of life. Digital mobile health (mHealth) techniques have the potential to aid the delivery of self-management interventions for COPD. We developed an mHealth intervention (Self-Management supported by Assistive, Rehabilitative, and Telehealth technologies-COPD [SMART-COPD]), delivered via a smartphone app and an activity tracker, to help people with COPD maintain (or increase) physical activity after undertaking pulmonary rehabilitation (PR).

Objective: This study aimed to determine the feasibility and acceptability of using the SMART-COPD intervention for the self-management of physical activity and to explore the feasibility of conducting a future randomized controlled trial (RCT) to investigate its effectiveness.

Methods: We conducted a randomized feasibility study. A total of 30 participants with COPD were randomly allocated to receive the SMART-COPD intervention (n=19) or control (n=11). Participants used SMART-COPD throughout PR and for 8 weeks afterward (ie, maintenance) to set physical activity goals and monitor their progress. Questionnaire-based and physical activity–based outcome measures were taken at baseline, the end of PR, and the end of maintenance. Participants, and health care professionals involved in PR delivery, were interviewed about their experiences with the technology.

Results: Overall, 47% (14/30) of participants withdrew from the study. Difficulty in using the technology was a common reason for withdrawal. Participants who completed the study had better baseline health and more prior experience with digital technology, compared with participants who withdrew. Participants who completed the study were generally positive about the technology and found it easy to use. Some participants felt their health had benefitted from using the technology and that it assisted them in achieving physical activity goals. Activity tracking and self-reporting were both found to be problematic as outcome measures of physical activity for this study. There was dissatisfaction among some control group members regarding their allocation.
Conclusions: mHealth shows promise in helping people with COPD self-manage their physical activity levels. mHealth interventions for COPD self-management may be more acceptable to people with prior experience of using digital technology and may be more beneficial if used at an earlier stage of COPD. Simplicity and usability were more important for engagement with the SMART-COPD intervention than personalization; therefore, the intervention should be simplified for future use. Future evaluation will require consideration of individual factors and their effect on mHealth efficacy and use; within-subject comparison of step count values; and an opportunity for control group participants to use the intervention if an RCT were to be carried out. Sample size calculations for a future evaluation would need to consider the high dropout rates.

(JMIR Mhealth Uhealth 2020;8(6):e16203) doi:10.2196/16203

KEYWORDS
mobile health; mHealth; chronic obstructive pulmonary disease; feasibility; physical activity; activity tracker; Fitbit; self-management; health behavior change; pulmonary rehabilitation

Introduction

Chronic Obstructive Pulmonary Disease

Chronic obstructive pulmonary disease (COPD) is one of the most prevalent chronic conditions (CCs), and one of the leading causes of death and disability, in the United Kingdom [1] and worldwide [2]. COPD is characterized by progressive and nonreversible narrowing or inflammation of the airways or alveoli in the lungs [3,4]. People with COPD experience symptoms such as breathlessness, frequent chest infections, reduced ability to exercise, and impaired day-to-day functioning [5-7]. COPD and its treatment cost the UK’s National Health Service (NHS) approximately 800 million pounds annually [3,5].

Self-Management

Even with appropriate medical care, people with COPD experience symptoms and functional challenges on a daily basis, and therefore, they must engage in long-term self-management to maintain their physical, social, and psychological health [8,9]. Teaching people with CCs to self-manage their health has become an important strategy for alleviating symptom burden and improving quality of life and is advocated within the NHS as a method of empowering people to take control of their health [10]. Participation in self-management activities requires changes to personal health behaviors by the individual with COPD. For example, COPD is associated with low physical activity levels [11]—higher levels of physical activity in the management of COPD are associated with a lower risk of hospital admission, a lower risk of COPD-related and all-cause mortality, and a higher health-related quality of life compared with those with lower activity levels [12,13].

Pulmonary Rehabilitation

Pulmonary rehabilitation (PR) is an intervention that aims to foster self-management among people with COPD in the United Kingdom. It is a group-based program that takes place over a minimum of 6 weeks and aims to teach people with COPD (and other lung conditions) to self-manage their condition [5,14]. UK National Institute for Health and Care Excellence 2018 guidelines [14] stipulate that the service should be available nationally for all people with COPD who have recently been hospitalized with the condition or who are functionally restricted by the condition. The program includes physical exercise of the upper and lower extremities, education about different aspects of the condition and how to manage them (eg, breathlessness), and strategies for improving daily functioning [4,5,14]. PR has demonstrated a number of benefits for people with COPD, including improved exercise capacity, alleviation of symptoms, reduced number and severity of exacerbations, reduced depression, and improved quality of life and sense of control [15]. In addition, PR has been demonstrated as cost-effective in the UK context [5]. However, long-term maintenance of increased physical activity and self-management behaviors after PR completion is a significant challenge [16,17].

Health Behavior Change

It is advantageous to design health behavior change interventions that are underpinned by behavior change theories or models, eg, the Behavior Change Wheel (BCW) [18]. Within the BCW model, intervention functions and policy categories are arranged around a central hub that outlines three different sources for health-related behaviors: opportunity, capability, and motivation. By providing education, training, persuasion, and environmental restructuring and enablement, PR can increase the social and physical opportunity for patients with COPD to engage in self-management, increase the physical and psychological capability of patients to engage in self-management, and help patients feel motivated to carry out self-management behaviors. The potential for digital technology to help people with COPD to self-manage their condition is increasingly being investigated [19,20] and advocated in the NHS [10,21]. Bartlett et al [22] demonstrated that people with COPD find support with a primary task, and dialog support (eg, through feedback), to be persuasive technological strategies to help increase their levels of physical activity. Therefore, a technological intervention for self-management that incorporates these elements could increase an individual’s capability, opportunity, and motivation to carry out that behavior (eg, motivation through encouraging feedback).

Mobile Health

Although the exact definition is disputed, mobile health (mHealth) broadly refers to medical or health care interventions delivered through mobile technology (eg, smartphones) [23]. Digital technologies such as mHealth offer several advantages over more traditional forms of care, including low up-front cost [24]; familiarity and convenience for patients [21]; better access to information [24]; improved communication between patients and health care professionals (HCPs) [21,24,25]; provision of real-time feedback to patients [24]; and allowing patients to
monitor their own data [24,25]—all of which potentially increase health service efficiency and reduce costs [21,25]. Although mHealth tools represent a promising means to encourage greater self-management of COPD, findings from a systematic review in this field were inconclusive owing to a high risk of bias in the included studies, thus indicating the need for more research [19]. Digital health interventions should be evidence based, person based, and robustly evaluated, with considerations given to future implementation of the intervention at an early stage of its development [20].

**Intervention Development**

We developed an mHealth-based self-management intervention for COPD. According to the Technology Acceptance Model, the perceived usefulness of a technological intervention directly affects an individual’s intention to use the technology in question [26]. Therefore, we carried out a large amount of exploratory work with the intended users and stakeholders of the technology. Qualitative semistructured interviews were conducted with people with COPD (n=15), their family members (n=5), and HCPs who work in PR services (n=7). During the interviews, we explored participants’ experiences of COPD self-management, their priorities for self-management, and their views on using digital technology to aid self-management. We also showed participants examples of ways in which digital technology could enhance capability, opportunity, and motivation [18] for self-management of COPD, eg, through goal setting and automatic monitoring of goals, demonstrating different types of devices and their functions, etc. The purpose of these interviews was to explore participants’ reactions to the possibility of using digital technology to help with self-management of the condition and to feed into the design of such an intervention.

During these exploratory interviews, both people with COPD and HCPs identified physical activity as being a high priority for COPD self-management. Feedback from the interviews informed the development of a prototype intervention that used an activity tracker and a smartphone app to help people with COPD set physical activity goals and monitor their progress. A total of 5 in-house researchers (unconnected with the project) and 5 participants with COPD from the exploratory interviews were later shown the prototype intervention and were asked to carry out think-aloud tasks using the technology, in accordance with a user-centered design [27]. The 5 participants with COPD involved in this stage of usability testing were selected to include a range of ages, gender, and COPD severities. The usability testing helped with assessing the intervention’s relevance and usability, identified problems in its operation, and helped further refine the intervention. After this stage, 2 people with COPD used the intervention over a period of many weeks. They relayed their experiences of using the intervention and offered suggestions for improvement.

The results of these interviews, of usability testing, and of a scoping literature review of existing best practice guidelines informed the development of a Self-Management supported by Assistive, Rehabilitative, and Telehealth technologies-COPD (SMART-COPD) app for COPD self-management and informed strategies for its use. In addition to the emphasis on physical activity, the importance of HCP support in the path to self-management was also emphasized. Therefore, the intervention was incorporated within the PR program, with the aim of encouraging individuals to maintain increased levels of physical activity after completing PR. In accordance with the BCW approach, the app provides motivation to self-manage physical activity through personalized feedback along with PR and provides the capability and opportunity to continue self-managing physical activity after PR. The development of the mHealth intervention is summarized in a short YouTube video [28].

**Feasibility and Acceptability**

Feasibility studies are carried out before large-scale studies (such as randomized controlled trials, RCTs) with the aim of establishing whether an intervention can be used and, if so, how it should be used [29]. The feasibility stage includes testing the intervention for its acceptability, estimating likely recruitment rates and retention of participants, and testing out design elements of a larger study [30]. A mixture of quantitative and qualitative methods is likely needed to establish feasibility [30].

An important element of the feasibility of carrying out a larger study is the acceptability of the intervention itself. In this study, we assessed the acceptability of the intervention using a combination of both qualitative (eg, interview data) and quantitative (eg, usage data collected by the intervention) data. Assessment of acceptability included attitude toward the intervention; burden of the technology; perceived effectiveness of the intervention; how well the intervention fits with the participants’ perceived value system; intervention coherence; and self-efficacy in being able to use the intervention [31]. Although objective data, such as dropout rates, provided a quantitative indication for ease of use of the intervention, the research team did not set predetermined thresholds for levels that would deem the intervention feasible or not feasible. The study evaluated the feasibility of delivering a complex intervention in a complex health care setting; therefore, qualitative interview feedback and reasons for withdrawal were necessary to understand the nuances behind the intervention’s feasibility.

**Research Questions**

This paper reports the results of a randomized feasibility study. The research questions (RQs) were as follows:

- **RQ1:** Is it feasible and acceptable to use the SMART-COPD intervention within PR to encourage people with COPD to maintain (or increase) their physical activity levels after PR?
- **RQ2:** Is it feasible to conduct a future large-scale RCT to investigate the effectiveness of the intervention?

Specifically, the following questions were addressed:

- How do people with COPD and HCPs react to the technology? What are their views on the technology, and on whether it is feasible to use the technology for physical activity in COPD? (related to RQ1)
- Is the technology acceptable to people with COPD and HCPs? For example, do they use the technology as...
intended? Do they find the technology easy to use? Which parts of the technology do they use? Are there any problems with the technology? (RQ1)

- What are the recruitment and dropout rates for the study? What do these patterns tell us about the acceptability of the technology and the feasibility of conducting a future RCT? (RQ1 and RQ2)
- Which outcome measures should be used for a larger-scale evaluation of the technology? (RQ2)
- How do people with COPD react to randomized assignment and to being in the Control group? (RQ2)
- How should the technology be deployed: both within health care services and within an RCT? Are any changes needed to increase the feasibility of using technology in this way? (RQ1 and RQ2)

The overall aim of the study was to determine the feasibility and acceptability of both the intervention and the possibility of carrying out a future RCT.

### Methods

#### The Self-Management Supported by Assistive, Rehabilitative, and Telehealth Technologies-Chronic Obstructive Pulmonary Disease Intervention

The SMART-COPD intervention is an Android smartphone mobile app used in conjunction with a Fitbit wearable activity tracking device. The app encourages physical activity through goal setting, self-monitoring, and feedback, based on behavior change and persuasive technology principles [18,22]. The app includes three elements of physical activity: general cumulative activity over each day (ie, step count); a timed daily walk; and timed daily exercises based on standard PR exercises (eg, leg lifts). The integrated content of the app is summarized in Table 1.

### Table 1. Summary of physical activity components within the SMART-COPD intervention.

<table>
<thead>
<tr>
<th>Facet</th>
<th>Step-count (activity tracker and app)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Daily walk (app only)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Exercises (app only)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal type</td>
<td>Number of steps (measured by activity tracker)</td>
<td>Length of walk in minutes (measured by phone’s accelerometer)</td>
<td>Length of time exercising (measured via manual timer on phone)</td>
</tr>
<tr>
<td>Feedback</td>
<td>Daily step-count visible on activity tracker, and feedback graphs on phone</td>
<td>App shows flower gaining petals as they get closer to their goal, and feedback graphs on phone</td>
<td>Videos demonstrating different exercises, timed doing exercises, and feedback graphs on phone</td>
</tr>
</tbody>
</table>

<sup>a</sup>Individualized goals set in partnership with PR staff. Participants self-monitor progress via described outlets.

<sup>b</sup>COPD: chronic obstructive pulmonary disease.

Participants decided on a personal starting goal for each of the three activities based on advice from their PR team and their own preferences. It was important that participants felt ownership of their goals and that both the participant and HCP felt they were achievable. The intention was that they should consistently achieve their activity goals on a daily basis and, if possible, gradually increase each physical activity goal over time. The app was used initially in conjunction with the PR program, with continued use once the PR program had finished.

Study participants were provided a Motorola smartphone with the SMART-COPD app installed and a Fitbit activity tracker for the duration of the study. Each smartphone was fitted with a SIM card to enable remote data transfer. The app could be used in conjunction with 1 of 3 Fitbit models: Charge HR (wrist-worn); Charge 2 (wrist-worn); or One (hip-worn). Fitbit activity trackers use low-energy Bluetooth to automatically transmit (or sync) data periodically whenever the device is in the proximity of a smartphone with an appropriately configured Fitbit app. These data are then transferred to Fitbit’s internet servers, where they can be accessed in the SMART-COPD app. The devices have a rechargeable battery with an approximately 5-day battery life. Fitbit One had previously demonstrated high accuracy compared with other low-cost activity trackers even for slower walking speeds (a factor highly relevant for COPD) [32]. The decision to use Fitbit as a step count device was further supported by in-house comparisons of step count accuracy of various activity monitoring devices (including Fitbit One and Charge HR), at different walking speeds, carried out by researchers and people with COPD.

#### The Setting

In total, PR teams at 3 NHS sites in Northern England, United Kingdom, participated in the feasibility study. Northern England has one of the highest rates of lung disease in the United Kingdom, possibly due to greater socioeconomic deprivation and higher rates of smoking compared with the South of England [3,4,33]. Preliminary work was conducted to map current PR care pathways (eg, workshops with PR staff; observation of PR sessions, numbers, and demographics of referred patients; etc), and we worked together with the PR teams to determine how the intervention might best be used within the PR program. Each PR service was delivered over a 6- to 7-week timeframe, with similar exercises and educational content. Therefore, the same study procedure was used across all sites.

#### Study Design

The study was a randomized feasibility study [34] using both quantitative and qualitative methods. The Medical Research Council Framework for the evaluation of complex interventions emphasizes the feasibility stage as a means of testing procedures...
and estimating parameters for a future large-scale evaluation [30].

**Participants**

The aim was to recruit 30 individuals who were formally diagnosed with COPD and who were attending PR in 1 of the 3 study sites. This sample size was chosen based on advice from an in-house statistician and on a study by Julious [35]. Potential participants met the inclusion criteria if they were attending PR. There were no exclusion criteria based on age, comorbidities, or having previously attended PR for managing COPD. Participants did not need any previous experience of using digital technology.

**Procedure**

Details of the study procedure are summarized in Figure 1. PR attendees were assessed for PR eligibility before starting the program. Participants attended PR twice weekly for 6 to 7 weeks. Each week PR physiotherapists informed the researchers if there were any new starters with COPD expected at PR sessions in the coming week. In their second PR session (during week 1), participants were asked if they would be happy for a researcher to speak with them about the feasibility study and what it would involve. Potential participants were given a participant information sheet and asked to contact the research team within the next week if they wished to take part.

**Figure 1.** Summary of participants’ progress through the study. COPD: Chronic Obstructive Pulmonary Disease; ISWT: Incremental Shuttle Walk Test; PR: pulmonary rehabilitation.

<table>
<thead>
<tr>
<th>TIMEFRAME</th>
<th>TASKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR ASSESSMENT</td>
<td>- Eligibility for PR</td>
</tr>
<tr>
<td></td>
<td>- Baseline ISWT</td>
</tr>
<tr>
<td>PR WEEK 1</td>
<td>New starters with COPD identified</td>
</tr>
<tr>
<td></td>
<td>- Researcher speaks to new starters about study</td>
</tr>
<tr>
<td></td>
<td>- New starter shows interest – baseline appointment made</td>
</tr>
<tr>
<td>PR WEEK 2</td>
<td>Baseline researcher visit</td>
</tr>
<tr>
<td></td>
<td>- Written informed consent</td>
</tr>
<tr>
<td></td>
<td>- Baseline outcome measures</td>
</tr>
<tr>
<td></td>
<td>- Randomized to intervention or control</td>
</tr>
<tr>
<td></td>
<td>- Shown how to use technology</td>
</tr>
<tr>
<td></td>
<td>- Begin using technology</td>
</tr>
<tr>
<td>PR WEEKS 3-5</td>
<td>- Continue attending PR</td>
</tr>
<tr>
<td></td>
<td>- Continue using technology</td>
</tr>
<tr>
<td>PR WEEK 6</td>
<td>- Finish PR</td>
</tr>
<tr>
<td></td>
<td>- Follow up 1 ISWT</td>
</tr>
<tr>
<td>MAINTENANCE WEEK 1</td>
<td>Follow-up 1 (F1) researcher visit</td>
</tr>
<tr>
<td></td>
<td>- F1 outcome measures</td>
</tr>
<tr>
<td>MAINTENANCE WEEKS 2-8</td>
<td>- Continue using technology</td>
</tr>
<tr>
<td>MAINTENANCE WEEKS 8-9</td>
<td>Follow-up 2 (F2) researcher visit</td>
</tr>
<tr>
<td></td>
<td>- F2 ISWT</td>
</tr>
<tr>
<td></td>
<td>- F2 outcome measures</td>
</tr>
<tr>
<td></td>
<td>- Qualitative interview</td>
</tr>
<tr>
<td></td>
<td>- Technology collected</td>
</tr>
</tbody>
</table>
Data Collection and Randomization

If a participant agreed to take part, then a baseline appointment was organized with one of the research team members for week 2 of PR. At the baseline visit, written informed consent was obtained and the participant completed a set of questionnaires (see Quantitative Outcome Measures). The participant was then randomized to 1 of 2 conditions:

- Group 1 (intervention) used the app and activity tracker to monitor, maintain, and (if possible) increase their physical activity during their time in PR (the PR phase) and for a further 8 weeks afterwards (the maintenance phase).
- Group 2 (control) wore a blinded activity tracker for the PR phase and maintenance phase. A strong black tape was used to cover the activity tracker’s screen so the participant would not be able to see their step count. This group was also provided with a smartphone so that data from the activity tracker could automatically be sent to, and stored on, the phone.

Uneven randomization was used, whereby two-thirds of participants were assigned to group 1 and one-third to group 2. This was due to the need to assess the acceptability and usability of the intervention. A blinded researcher used sealed opaque envelopes to generate randomization. This method of allocation was chosen because it mimics the more rigorous software-generated randomization method used in RCTs, in a manner that was satisfactory for the purposes of the feasibility study.

After allocation, the researcher demonstrated the technology to the participant and set up baseline activity goals within the app. Participants were provided with an instruction manual and the research team’s contact details.

Participants then used the app and activity tracker, or wore a blinded activity tracker, during the PR phase in accordance with their allocation. An appointment was made for a member of the research team to visit the participant at the end of PR and take follow-up 1 (F1) measurements. At this point, the set of questionnaires applied at baseline was repeated.

Another appointment was made for 8 weeks after finishing PR (follow-up 2 or F2), ie, the end of maintenance. If possible, participants were seen at the PR facility so that a final incremental shuttle walk test (ISWT) could be conducted (see Quantitative Outcome Measures), a final set of questionnaires could be completed, and a semistructured qualitative interview could be conducted about their experiences on the study. The interview explored their experiences of using the technology, perceived benefits and barriers to using the technology, acceptability of randomization, perceived impact of the technology on physical activity, and tolerability of outcome measures (see Multimedia Appendix 1). During the final ISWT, participants wore a Fitbit Charge 2 (wrist-worn), a Fitbit One (hip-worn), and Axivity motion sensors on their hip and wrist to compare step count accuracy across different devices for individual participants. The devices were found to be comparable in accuracy, although detailed results of this comparison are not reported here. Finally, the research team retrieved the technology from the participant.

Clinical support (via the relevant PR team during PR sessions) and technical support (via the research team) was available to all participants throughout their time in the study. The protocol and all materials used for the study were reviewed by people with COPD for relevance and comprehensibility.

Health Care Professionals

At the end of the recruitment period, PR team members (eg, respiratory physiotherapists) were contacted via email with a participant information sheet attached and were asked to contact the research team if they wished to take part in a qualitative semistructured interview or focus group discussion. The purpose was to explore HCPs’ opinions of using the technology alongside PR, including perceived benefits and barriers to using the technology, acceptability of the technology, and perceived impact on participants’ physical activity (see Multimedia Appendix 2).

Quantitative Outcome Measures

The feasibility of using a number of quantitative outcome measures (for a future RCT) was assessed during the study, eg, for relevance and ease of completion. The System Usability Scale (SUS) [36] was also included as an objective measure of ease of use of the intervention. The quantitative outcome measures used are summarized in Table 2.

Exercise capacity was assessed using the ISWT, a well-validated measure [37] used as part of standard clinical practice by all three PR sites. Patients completed a baseline ISWT at their assessment visit and another ISWT in their final week of PR (ie, at F1) as part of standard practice. The research team requested these scores for all participants in the study and added an additional ISWT for participants at the end of the maintenance phase (F2). This was conducted at the PR facility by a respiratory physiotherapist. The research team collected all other outcome measures.

https://mhealth.jmir.org/2020/6/e16203
Table 2. Summary of measures taken from participants at different time points.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Baseline</th>
<th>Follow-up 1</th>
<th>Follow-up 2</th>
<th>Continuous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, gender, ethnicity, postcode, medical conditions, previous PR, and previous experience with technology</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Medical Research Council Breathlessness Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COPD severity: 1 (“not troubled by breathlessness except on strenuous exercise”) to 5 (“too breathless to leave the house, or breathless when undressing”) [38]</td>
<td>X</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical activity: step count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>As measured by an activity tracker</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>X</td>
</tr>
<tr>
<td>Physical activity: CHAMPS [d] questionnaire</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAMPS [39]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Exercise capacity: ISWT [e]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ISWT [37]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Functioning and quality of life: SGRQ [f]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGRQ [40]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Anxiety and depression: PHQ-9 [g]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHQ-9 [41]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Exercise self-efficacy: Ex-SRES [h]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex-SRES [42]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Symptoms: CAT [i]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT [43]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Cost-effectiveness: EQ-5D-3L [j]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ-5D-3L [44] for a future cost-effectiveness assessment</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Usability: SUS [k]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUS [36] to assess the usability of technology</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
</tbody>
</table>

PR: pulmonary rehabilitation.
X: Measure taken at indicated timepoint.
N/A: not applicable.
CHAMPS: Community Healthy Activities Model Program for Seniors.
ISWT: incremental shuttle walk test.
SGRQ: St George's Respiratory Questionnaire.
PHQ: Patient Health Questionnaire.
Ex-SRES: Exercise Self-Regulatory Efficacy Scale.
CAT: Chronic Obstructive Pulmonary Disease Assessment Test.
EQ-5D-3L: EuroQol 5 Dimensions 3 Level
SUS: System Usability Scale.

Quantitative Analysis
Descriptive statistics were calculated for demographic data, ISWT scores, and questionnaire-based outcome measures in SPSS. Data collected via the SMART-COPD app on amount and types of physical activity were summarized and explored in Microsoft Excel to provide insights into how the intervention was used.

Qualitative Analysis
Qualitative interviews with patients with COPD and HCPs were transcribed verbatim, and a thematic analysis [45] was used to identify key themes within the data (using NVivo software). The first author explored the transcript data, taking notes on pre-existing and emerging themes. These notes were used to build a coding framework, which was used to code the transcript data within NVivo. All data within individual themes were then explored and summarized (including where there were
differences between participants), to produce summaries of each theme and also to identify relationships between themes.

**Ethics**

The study received NHS research ethics approval (15-YH-0458), as well as Health Research Authority and research governance approval from each NHS site. The feasibility study was registered on a clinical trials database (NCT02691104).

### Results

**Recruitment and Dropout Rates**

A total of 30 people with COPD participated in the feasibility study: 19 participants were assigned to the intervention group and 11 were assigned to the control. 16 participants completed all three data collection points and 14 participants withdrew from the study. The groups were well matched on most demographics. Participants’ demographics are summarized in Table 3.

#### Table 3. Summary of participants’ baseline demographics and measurements.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Intervention</th>
<th>Control</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>68.0 (63.0-72.0)</td>
<td>66.0 (60.0-70.0)</td>
<td>67.5 (60.0-70.5)</td>
</tr>
<tr>
<td>Range</td>
<td>45-75</td>
<td>33-75</td>
<td>45-75</td>
</tr>
<tr>
<td>Gender (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Female</td>
<td>11</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Medical Research Council Breathlessness score(^a) (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>PR(^b) attendances (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First time</td>
<td>11</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>Been before</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Ethnicity (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White British</td>
<td>19</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>“Regularly use computer” (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>No</td>
<td>11</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>“Regularly use mobile phone” (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>17</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>“Regularly use tablet” (frequency)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>No</td>
<td>11</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

\(^a\)1=least severe and 5=most severe.
\(^b\)PR: pulmonary rehabilitation.

**Study Withdrawal**

In the intervention group, 47% (9/19) of participants withdrew, and in the control group 46% (5/11) of participants withdrew; thus, attrition rates were similar for both groups. In total, 12 participants withdrew before F1: 3 participants withdrew either at baseline or within 2 weeks of baseline; and 9 participants withdrew within 3 to 6 weeks into the study. A further 2 participants withdrew between F1 and F2. More women withdrew compared with men (9/17, 53% vs 5/13, 38%, respectively). Withdrawers also had a lower median age compared with completers (median 65.5, IQR 59.8-70.5 years vs median 68.0, IQR 61.0-71.3 years, respectively). The most common (voluntarily given) reasons for withdrawing included ill health (n=5); withdrew from PR (n=4); burden of research, of technology, or of completing daily exercises (n=4); technical issues or frustrations with the technology (n=4); and disappointment at having been assigned to the control group (n=3). Some withdrawers cited more than one of these reasons.
Several withdrawers (n=3) still liked the concept of the app, and 2 participants who had stopped attending PR would have liked to continue using the app.

Descriptive statistics on baseline outcome measures revealed a pattern of differences between participants who completed the study and those who withdrew. Withdrawers showed signs of having worse baseline disease severity and health compared with those who completed the study. These comparisons are summarized in Table 4.

Participants who withdrew had worse baseline scores on exercise capacity, quality of life, and depression compared with those who completed. However, sample sizes were small, so this finding should be interpreted with caution.
Table 4. Comparison of completed vs withdrawn participants on baseline outcome measures.

<table>
<thead>
<tr>
<th>Baseline measures</th>
<th>Completed</th>
<th>Withdrawn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline ISWT score (meters)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>255.0 (172.5-332.5)</td>
<td>170.0 (105.0-305.0)</td>
</tr>
<tr>
<td>Range</td>
<td>90-550</td>
<td>40-490</td>
</tr>
<tr>
<td><strong>Baseline SGRQ score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>52.0 (44.2-63.7)</td>
<td>62.4 (52.3-72.3)</td>
</tr>
<tr>
<td>Range</td>
<td>32.3-78.1</td>
<td>42.3-79.6</td>
</tr>
<tr>
<td><strong>Baseline SGRQ current health question</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Very poor</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Poor</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Fair</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Good</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Very good</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Baseline CHAMPS score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>2017.3 (1220.5-5237.3)</td>
<td>2468.5 (1193.8-3138.7)</td>
</tr>
<tr>
<td>Range</td>
<td>68.0-968.3</td>
<td>437.7-9644.3</td>
</tr>
<tr>
<td><strong>Baseline Ex-SRES score (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>54.1 (27.7-72.0)</td>
<td>54.1 (32.4-69.7)</td>
</tr>
<tr>
<td>Range</td>
<td>10.6-93.1</td>
<td>13.1-86.3</td>
</tr>
<tr>
<td><strong>Baseline PHQ-9 score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>7.0 (4.3-12.5)</td>
<td>10.0 (5.5-15.0)</td>
</tr>
<tr>
<td>Range</td>
<td>2.0-16.0</td>
<td>2.0-21.0</td>
</tr>
<tr>
<td><strong>Baseline CAT score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>23.0 (16.5-25.0)</td>
<td>23.0 (16.8-26.8)</td>
</tr>
<tr>
<td>Range</td>
<td>13.0-32.0</td>
<td>12.0-37.0</td>
</tr>
<tr>
<td><strong>EQ-5D score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>8.0 (7.0-9.0)</td>
<td>8.0 (8.0-10.3)</td>
</tr>
<tr>
<td>Range</td>
<td>5.0-11.0</td>
<td>6.0-12.0</td>
</tr>
<tr>
<td><strong>EQ-5D scale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>60.0 (50.0-70.0)</td>
<td>50.0 (40.8-62.5)</td>
</tr>
<tr>
<td>Range</td>
<td>28.0-90.0</td>
<td>20.0-90.0</td>
</tr>
</tbody>
</table>

aISWT: incremental shuttle walk test; higher score=greater distance walked.
bSGRQ: St George’s Respiratory Questionnaire; lower score=better quality of life.

cCHAMPS: Community Healthy Activities Model Program for Seniors; higher score=more physical activity.

dEx-SRES: Exercise Self-Regulatory Efficacy Scale; higher score=more exercise-related self-efficacy.

ePHQ-9: Patient Health Questionnaire for Depression; higher score=more depressed.

fCAT: Chronic Obstructive Pulmonary Disease Assessment Test; higher score=more COPD symptoms.

gEQ-5D-3L: EuroQol 5 Dimensions 3 Level score; higher score=worse health.

hEQ-5D-3L scale: EuroQol 5 Dimensions 3 Level scale; scored 0-100, where 100=best health they can imagine.

Use of the Technology

For intervention participants, physical activity and usage data collected by the SMART-COPD app were transmitted by email using the smartphone’s mobile data connection. These data provided indications of how the app was used and how often. Table 5 summarizes the percentage of days on which each physical activity component of the app was used by each participant.

Table 5. Individual participants’ use of different features of the SMART-COPD app.

<table>
<thead>
<tr>
<th>Participant</th>
<th>App use, no. days (%)</th>
<th>Steps recorded(^a), no. days (%)</th>
<th>Exercise recorded, no. days (%)</th>
<th>Daily walk recorded, no. days (%)</th>
<th>Total days participant had intervention, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1(^b)</td>
<td>12 (32)</td>
<td>15 (41)</td>
<td>6 (16)</td>
<td>9 (24)</td>
<td>37</td>
</tr>
<tr>
<td>2(^b)</td>
<td>18 (64)</td>
<td>25 (89)</td>
<td>3 (11)</td>
<td>3 (11)</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>108 (90.0)</td>
<td>107 (89.2)</td>
<td>101 (84.2)</td>
<td>14 (11.7)</td>
<td>120</td>
</tr>
<tr>
<td>6</td>
<td>53 (51.0)</td>
<td>53 (51.0)</td>
<td>36 (34.6)</td>
<td>19 (18.3)</td>
<td>104</td>
</tr>
<tr>
<td>8</td>
<td>85 (98)</td>
<td>86 (99)</td>
<td>64 (74)</td>
<td>65 (75)</td>
<td>87</td>
</tr>
<tr>
<td>11(^b)</td>
<td>23 (100)</td>
<td>18 (78)</td>
<td>1 (4)</td>
<td>10 (43)</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>85 (94)</td>
<td>83 (92)</td>
<td>57 (63)</td>
<td>79 (88)</td>
<td>90</td>
</tr>
<tr>
<td>15</td>
<td>62 (84)</td>
<td>66 (89)</td>
<td>1 (1)</td>
<td>56 (76)</td>
<td>74</td>
</tr>
<tr>
<td>19</td>
<td>100 (89.3)</td>
<td>99 (88.4)</td>
<td>41 (36.6)</td>
<td>41 (36.6)</td>
<td>112</td>
</tr>
<tr>
<td>21</td>
<td>19 (20)</td>
<td>79 (82)</td>
<td>8 (8)</td>
<td>0 (0)</td>
<td>96</td>
</tr>
<tr>
<td>22(^b)</td>
<td>69 (100)</td>
<td>68 (99)</td>
<td>4 (6)</td>
<td>58 (84)</td>
<td>69</td>
</tr>
<tr>
<td>23(^b)</td>
<td>6 (100)</td>
<td>6 (100)</td>
<td>3 (50)</td>
<td>5 (83)</td>
<td>6</td>
</tr>
<tr>
<td>24</td>
<td>19 (17)</td>
<td>31 (28)</td>
<td>5 (5)</td>
<td>15 (14)</td>
<td>111</td>
</tr>
<tr>
<td>27</td>
<td>71 (63)</td>
<td>88 (79)</td>
<td>7 (6)</td>
<td>20 (18)</td>
<td>112</td>
</tr>
<tr>
<td>30</td>
<td>94 (96)</td>
<td>13 (13)</td>
<td>5 (5)</td>
<td>83 (85)</td>
<td>98</td>
</tr>
<tr>
<td>Mean</td>
<td>55 (71)</td>
<td>56 (72)</td>
<td>23 (29)</td>
<td>32 (41)</td>
<td>78</td>
</tr>
</tbody>
</table>

\(^a\)Steps were recorded by an activity tracker even if the Self-Management supported by Assistive, Rehabilitative, and Telehealth technologies-Chronic Obstructive Pulmonary Disease app itself was not used.

\(^b\)Participants who later withdrew from the study.

A total of 3 intervention participants withdrew immediately or within 1 to 2 weeks of receiving the technology. No app use data were recorded for these participants. One intervention participant (17) withdrew 5 weeks into the study but had no app use data recorded. Technical issues were noted for this participant (eg, the activity tracker not holding its charge); however, it is unclear if this was the reason for the lack of data from this participant.

On average, the SMART-COPD app was used on 73% of days on which it was deployed to a participant, although individual participants’ usage patterns varied widely. The steps component of the intervention was the most frequently used physical activity strategy overall. However, use of the steps component decreased for some participants when they moved into maintenance. Participants generally maintained consistent (high or low) usage levels for daily walks and exercises across the PR and maintenance phases.

Outcome Measure: Step Counts

Participants did not show a consistent pattern overall with respect to whether their step counts increased, decreased, or stayed the same over time. Most intervention participants who completed the study had a near-complete dataset for the full study timeframe. Only 2 control participants had near-complete datasets. Some gaps in the data were explained by technical issues or by participant illness, but many gaps were unexplained.

Outcome Measure: Incremental Shuttle Walk Test

The ISWT was found to be relevant and appropriate as an outcome measure for exercise capacity, as this was routinely collected in the three PR sites and participants were used to
completing it. However, the logistics of carrying out a third ISWT at F2 (when participants had already been discharged from the PR service) at times proved challenging owing to the need to coordinate availability between participants, the research team, and the PR team.

**Outcome Measure: Questionnaires**

All participants were generally satisfied with the questionnaires and length of time needed to complete them. Some participants requested help from the researcher to complete them, eg, asking the researcher to read questions aloud and complete answers on their behalf. Some questions felt repetitive to participants if a similar question was included on more than one questionnaire. However, the only questionnaire that caused significant problems in completion was the Community Healthy Activities Model Program for Seniors (CHAMPS) self-reported physical activity questionnaire. Participants did not identify with the Americanized nature of included activities or the wording of some of the questions (eg, shooting pool), which were felt to be less relevant to a British population. Researchers also noted problems in participants’ understanding of the timeframe to which the questions applied and confusion over calculating how often, or for how long, each activity was completed.

**System Usability Scale**

Individual participants’ SUS scores at F1 and F2 are summarized in Table 6 (for intervention participants who completed the study). Scores are out of 100, with a higher score indicating greater usability.

SUS scores were generally high compared with the industry standard average score of 68 [42], indicating the intervention had a higher than average usability level across technological systems from multiple industries. Participants’ SUS scores generally (but not always) increased with time spent using the technology.

**Table 6.** Individual intervention participants’ System Usability Scale scores at Follow-up 1 and Follow-up 2.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Follow-up 1</th>
<th>Follow-up 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>97.5</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>80</td>
<td>92.5</td>
</tr>
<tr>
<td>8</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>14</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>15</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>19</td>
<td>77.5</td>
<td>82.5</td>
</tr>
<tr>
<td>21</td>
<td>65</td>
<td>5</td>
</tr>
<tr>
<td>24</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>27</td>
<td>67.5</td>
<td>97.5</td>
</tr>
<tr>
<td>30</td>
<td>Missing</td>
<td>Missing</td>
</tr>
</tbody>
</table>

**Qualitative Results**

**People With Chronic Obstructive Pulmonary Disease: Reactions to the Intervention**

The interviews captured the views of 16 participants who completed the study (n=10 intervention group; n=6 control group). Qualitative data relating to the intervention were categorized into six main themes: technology; technical issues; previous experience (with digital technology); integration with PR; control group issues; and involvement in the project.

**Technology**

The intervention was generally well liked and well accepted among participants who completed the study. Most of these participants found the technology easy to use and were able to incorporate it into their daily routine, eg, putting on the activity tracker in the morning:

*Yes, absolutely no problem, once I’d got, had a shower and you know put it on and just carried on what I was doing… even forgetting sometimes I’d got it on.* [Patient 28]

Around half of (completed) the participants felt they had benefited (physically or psychologically) through enhanced confidence, monitoring of physical activity, and incentivization to exercise:

*If I hadn’t found this technology erm I would probably be just sat at home watching TV… it would never occur to me to do exercise at home and to move.* [Patient 3]

Many participants found the technology motivational. Feedback on activity and goal attainment motivated them to do more:

*You look at it in the day and… see how many steps you’ve done and it encourages you I think.* [Patient 6]

Two participants did not use the intervention in this way—they were already active and used the intervention to monitor activity that they were already doing rather than trying to increase it:

*It’s just the way I live. I’ve not done anything different to what I normally do… made me realise how much I was doing, or how little I was doing but I wouldn’t say it increased what I put into it.* [Patient 15]

One participant (21) disliked the smartphone and SMART-COPD app and found them difficult to use (supported...
by her F2 SUS score, although she was happy to continue wearing the activity tracker:

Rubbish cause I couldn’t do it, couldn’t do it, no way could I do it... I’ve never been able to use one of them phones for a start. [Patient 21]

Participants rarely updated goals. Reasons included conforming to what the experts (physiotherapists or researchers) recommended (and the physiotherapists did not explicitly tell them to increase these goals) and keeping the goal achievable:

I didn’t want to do it too much because I knew I’d be disappointed if I didn’t reach the goal I’d set. [Patient 3]

Opinions of the daily walk and exercise sections of the app were mixed. Even where the exercise section was used, the videos were not well liked—they were viewed as repetitive or participants were put off by people in the videos they perceived as less able than themselves:

I haven’t used the exercise bit because I looked at the exercise thing and it’s all people sat in chairs and looks like they’re all in old people’s homes, I’m a little bit more active than that. [Patient 15]

Reasons for low use of the daily walk section included difficulty getting the smartphone to register the walk and not wanting to walk outside alone:

I don’t go for daily walks so I thought I’ll give it a try in the house and I had to walk at such a speed [for the walk to register via the accelerometers in the smartphone] I was frightened of banging into doors and all sorts so in the end I decided not to do the daily walks. [Patient 3]

There were misunderstandings around the daily walk section of the app, with a few participants referring to the step count and daily walk sections interchangeably. At least two participants did not understand that the phone needed to register continuous movement and hence needed to be placed in the hand or the pocket to record the duration of the walk:

I can have walked for twenty minutes at the rehab centre on the err treadmill and it doesn’t show it, it never works... I have on the machine you know I stand it up on the machine. [Patient 6]

Many participants, from both groups, discovered the Fitbit app on the smartphone, even though it was not intended that either group should use it directly. Installation of the app was necessary for data synchronization but could be discovered by participants with an experience of, or curiosity about, using smartphones. Several used this, rather than the SMART-COPD app, to monitor their step counts. At least two control participants used the Fitbit app to monitor their steps and increase their activity despite it being against the planned intervention:

Just at odd times I did [check steps on Fitbit app], yes, yes just odd times, I didn’t cheat at all. [Patient 28, control]

Overall, the activity tracker was the most liked component of the intervention, and a few participants expressed a preference for using the activity tracker alone and leaving out the SMART-COPD app:

With Fitbit bit, chuck phone away and just leave Fitbit. As long as you’ve got the app that you can have on your own phone. [Patient 24]

Participants who already owned smartphones often expressed a preference for having the SMART-COPD app installed on their own phone rather than carrying and charging extra equipment. In total, 3 participants were conscious of not wanting to damage or lose the study’s equipment, which would be less of a worry if they were using their own technology:

The only thing that did affect me was the fear of losing it all... it’s using somebody else’s equipment and being responsible for it. [Patient 6]

Most participants did not carry the smartphone with them when they left the house. Participants were also less likely to use the technology if they were unwell, busy, or on holiday:

I just want to get up and go. If I’m taking my dog out for a walk I’ve got enough trouble getting leads, making sure there’s bags on it and harness on dog and treats to make sure they come back without having to take [the phone]. [Patient 24]

**Technical Issues**

The Fitbit device frequently stopped synchronizing with the Fitbit app. This happened for at least six participants. These occurrences were sometimes due to the device not being charged and, in one case, a participant accidentally deleting the Fitbit app from the smartphone. However, in most cases, these communication errors were unexplained:

They weren’t getting no data through and I said well I don’t know whether Fitbit what’s not charging or phone what’s not charging. [Patient 25]

There were a number of reports of itchiness, rash, or discomfort from wearing the activity tracker, particularly in warm weather:

I bruise very easily, my skin is wafer thin and it was very uncomfortable and it marked both arms so it had to go. [Patient 3]

A few participants who experienced early problems were offered the hip-worn Fitbit One as an alternative and seemed to get on well with this option. A few other participants expressed a preference for a hip-worn activity tracker at the interview stage, eg, for discretion.

There were at least three cases of the activity tracker catching on clothes and falling off:

I lost it in the churchyard once; luckily it was still there when I went back... it had caught on my sleeve and just pulled off. [Patient 15]

**Previous Experience**

Many participants had previous experience with using digital technology, such as smartphones and computers, and a few had a prior interest in technology:
[Talking about the game Pokémon Go] So I quite enjoy that and I’ve starting using my eggs as distances on how much I’ve walked. [Patient 24]

However, some participants had little or no previous experience, and these participants were more likely to encounter difficulties with using the intervention:

I don’t know how they work and I’m not interested to be honest. The learning how to work it, in my opinion I’m at wrong time, wrong side of life now to start worrying about technology. [Patient 25]

No one had previously used a Fitbit. Some participants knew about Fitbit devices through advertisements or family members, but had perceived them as being irrelevant for themselves or as being only for athletes:

I wouldn’t have known about that Fitbit cos I weren’t interested in things like that... I never took no notice. I always thought people did it when they went in gym, you know like they bought a Fitbit just to cycle and things like that. [Patient 29]

Some participants felt the intervention was less suited to people with COPD who were older, less technologically savvy, with more severe disease, or living alone, despite the app being perceived as easy to use:

I could understand how some people that are a bit older than me would struggle with it if they hadn’t got any support in the house. I mean I’ve got a husband and he’s pretty good with technical things but I didn’t have to really ask him, I figured it myself. [Patient 6]

Integration With Pulmonary Rehabilitation

There were few reported issues with the intervention being used alongside PR, and many participants liked their concurrent use. Most participants did not speak to the PR team about the technology. However, where they did speak to physiotherapists, they often found the HCP did not know much about the technology or the study (though they still showed an interest and tried to help with any issues):

They didn’t know how it worked. They didn’t, you know because I did ask at the beginning I was a bit flummoxed with it all erm and I did ask the physio that was there then and she, she had a look but she couldn’t tell me, but I figured it out myself in the end. [Patient 6]

Control Group

Most control group participants would have liked to see and use the intervention but were usually happy to take part. A few intervention participants felt they would have been disappointed to receive the control condition but stated they would have continued with the study:

Well, I honestly felt a bit disappointed... because erm I did want to experience trialling erm the Fitband [sic] as it should be used... but I realise the importance of that um, you need both aspects of it, so, I was still happy to do it. [Patient 07, control]

One control group participant who had no prior experience with digital technology stated he would have withdrawn if he had been allocated to the intervention group:

I’d have probably put it back in box and have rang em back up and said I don’t want to do it, it’s too much for me this, I’m not into it. [Patient 25]

When technical issues occurred, control group participants were less likely to realize that there was a technical issue compared with intervention group participants, as they could not see the Fitbit display or the Fitbit app (unless they had discovered it). When participants did realize that there was a technical issue, it was usually because they were using the Fitbit app or noticed the activity tracker no longer had a flashing light underneath:

Well part of it was when it wasn’t working, which I didn’t know, when I was unaware of that. [Patient 28]

One control group participant had been given a Fitbit Charge 2, whose charging cradle differs from the Charge HR. For this participant, the black tape covering the screen interfered with the charging cradle, which caused weeks of problems getting data from this participant (until the root cause was discovered).

Involvement in the Project

Participants did not generally contact the research team if they were having problems using the technology or were experiencing technical issues. These problems were usually detected when members of the research team called participants to arrange data collection visits. Some participants had lost the research team’s contact details; others seemed reluctant to bother researchers with issues they feared were due to their own incompetence or inexperience with technology:

No I’ve never contacted you. I was thinking about it in the beginning because I wasn’t getting, as I say I was a bit flummoxed by it all, but I figured it out so I didn’t. I’ve not phoned up or anything, I’ve not had any contact other than the visits. [Patient 6]

One participant who was initially resistant to using technology decided to try it after hearing the researcher’s explanation of the study:

Well when you when they first asked me, would I, you know at pulmonary rehab, would I take one and I was thinking no I can’t be doing with that, you know. Can’t be doing with that... because I didn’t understand it... until it was explained and then I thought, yes I’m going to have that, yes. [Patient 14]

Participants enjoyed taking part in the study and were mostly happy with their contact with the research team. Some participants had purchased, or planned to purchase, their own Fitbit to continue using after the study had finished:

It would never have entered my head to go and buy something to improve my condition, never, erm but now I’ve got another Fitbit waiting for me when I go home. [Patient 03]

Health Care Professionals: Reactions to the Technology

A total of 5 HCPs took part in either a focus group discussion or an interview exploring their experiences of participating in
the study. Five themes relevant to HCPs’ experiences with the technology are discussed: technology; recruitment; communication; workload; and suggested improvements.

Technology
One physiotherapist (with personal experience of using activity trackers) described the SMART-COPD app as clunky and visually unattractive. The staff felt some participants were motivated by the technology to achieve physical activity goals between PR sessions, although they also noted that some participants are naturally more motivated regardless of whether they have technology.

Recruitment
The staff reported that some participants got on with the technology better than others, and that it was usually (but not always) younger and more technologically experienced participants who adapted quickly and gained the most benefit. However, 2 staff members pointed out that older people are becoming more technologically experienced as the years progress, so this may not be an issue for future generations.

The study did not have any inclusion criteria around previous experience with technology. However, there were hints that a small number of potential participants may not have been put forward to the research team if staff members felt they would not benefit from the intervention, eg, if they had never used digital technology or did not seem motivated to benefit from PR. The staff also heard a few control group participants expressing disappointment with their allocation, eg, wondering about the purpose of their involvement.

Communication
During the feasibility study preparation, the research team conducted workshops with PR staff and involved them in planning the logistics of the intervention and the study. Unfortunately, it was (understandably) difficult to speak with the entire PR team; therefore, not all staff were briefed in-person on the study and technology, and key information did not always filter through to the entire team. In addition, physiotherapists within PR teams are frequently rotated to different locations. In the time it took to complete development work and get ethical amendments approved, some PR team members had changed, and even those who were involved in earlier stages of the study did not always recall how the technology or the study worked.

This was reflected in the staff interviews, in which physiotherapists often reported that they did not know much about the technology or the wider study. Owing to their own experiences with technology, PR team members were more able to help with generic smartphone issues but were not usually experienced with using activity trackers or the SMART-COPD app.

Workload
The study did not have a large impact on workload, as PR physiotherapists are accustomed to speaking with individual patients during PR sessions and checking their progress. One difficulty, however, was that the research team was not always told in good time when a new starter would be attending a PR session, especially when this was decided at short notice or when this coincided with staff absence on either side. Physiotherapists sometimes reported difficulties deciding appropriate physical activity goals for participants, as they did not think in terms of the number of daily steps a person with COPD should aim to achieve. There were also difficulties organizing appointments for conducting F2 ISWT tests (which were not a part of normal service delivery). Overall, it was felt that the SMART-COPD intervention had become a tool used alongside PR rather than incorporated within PR, although it was thought in some cases to have enhanced people’s PR experience and the benefits they gained from it.

Suggested Improvements
Suggestions for improving the study included a dedicated central email list to improve communication and sharing of responsibility for appointments and tasks. One HCP suggested asking participants to wear a blinded activity tracker for a week before starting the study to gain a better baseline physical activity level. This would help with setting appropriate physical activity goals for the individual and would help with working out whether the intervention was effective for that individual.

Discussion

Principal Findings
This randomized feasibility study examined the feasibility and acceptability of using a wearable activity tracker and the SMART-COPD app both within and following PR to encourage people with COPD to increase, or at least maintain, their physical activity levels (RQ1). The study also explored the feasibility of conducting a future RCT to investigate the effectiveness of the intervention (RQ2). The intervention shows potential in helping a subset of people with COPD to achieve physical activity goals. However, both the intervention and methods used would need to be modified if a future RCT were to be conducted.

Acceptability of the Intervention
A total of 30 people with COPD were recruited to the study. The 16 participants who completed the study were generally positive about the intervention (or liked the concept if they were in the control group). Some believed they had gained tangible benefits from using the intervention, both in terms of motivation to achieve physical activity goals and subsequent benefits to their physical or psychological health. Most participants who completed the study found the technology easy to use and most experienced no problems incorporating it into their daily lives. However, almost half of the participants (n=14) withdrew from the study. Some participants seemed to find the prospect of using digital technology for physical activity monitoring to be daunting or overwhelming. This was supported by the reasons given for withdrawing, and qualitative feedback from both patients and staff, which indicated that people who were older or less experienced with digital technology or did not have support at home might be wary of using this type of intervention. Overall, people who withdrew from the study had worse baseline scores on exercise capacity, quality of life, and depression compared with those who completed the study, which indicates that people with better COPD-related health may gain more.
benefits from mHealth-based interventions. This claim must however be interpreted with caution due to the small sample size.

Although mHealth is a promising intervention for the self-management of COPD, the evidence base around mHealth is currently mixed and underdeveloped [46-48]. In this study, there was arguably a dichotomy between people who completed the study and gained perceived physical and psychological benefits from the intervention and those who did not easily adopt the technology or found it to be a burden. This has been found for related interventions such as telehealth [46,49,50] and supports McCabe et al’s [19] hypothesis that patients with COPD with greater interest in technology may gain greater benefits from mHealth interventions. In addition, a recent literature review of wearable technologies for physical activity in COPD identified only a small number of RCTs with highly heterogeneous technologies and study designs, meaning no conclusions could be drawn about their effectiveness [51]. More evidence is needed on the use of wearables in COPD, along with an improved ability for accurate step count detection, and more robust guidelines are needed for clinical staff to implement wearable technologies for COPD [51].

The SMART-COPD app itself mostly worked well. However, one of the most challenging technical issues was the failure of activity data to be transmitted from the activity tracker to its corresponding smartphone app. This was frustrating for all concerned as it meant some participants’ physical activity data were not recorded, and indeed technical issues were cited by several participants as a reason for withdrawing. The wrist-strap also caused discomfort and could catch on clothes and fall off. These issues would need to be resolved if the intervention were to be used on a wider scale.

Although the intervention seemed acceptable as a tool for use alongside standard PR, a full integration of the intervention within service delivery was problematic. Communication errors occurred between the research team and PR team despite the best efforts of everyone involved. In addition, the logistical difficulties of briefing entire PR teams, coupled with the tendency of PR physiotherapists to be moved between teams, resulted in some PR team members feeling uninformed about the technology and the wider study. However, PR staff were still supportive of the intervention, especially when participants within their service told them about their positive experiences.

There are potential patient-level advantages to incorporating the SMART-COPD intervention within PR, eg, access to clinical support when first using the app. However, it is worth noting that a large proportion of patients with COPD referred to PR do not attend or do not complete the course [52,53]. Reasons for nonattendance and noncompletion include perceived lack of benefit, disruption to usual routine, poor access to transport, greater disease severity, lower quality of life, and greater symptoms of depression [53,54]. Even if the intervention were deployed within PR in the future, a large number of people with COPD would not have the opportunity to use this method of self-managing their physical activity. Therefore, future research should explore other ways of delivering mHealth interventions to people with COPD who do not access PR.

Feasibility of the Study Design

Almost half of recruited participants dropped out of the study. This finding has implications both for the acceptability of the intervention and for the sample size of a future RCT, which would need to account for high dropout rates. It could be argued that the inclusion criteria should be modified to target people with COPD who are deemed more likely to engage with this type of intervention. However, 1 or 2 individuals readily adopted the technology against their own or the physiotherapists’ expectations. HCPs in this study made the point that future generations coming through PR are likely to have more experience with digital technology and might therefore be more likely to engage with mHealth. These points suggest that the inclusion criteria should not exclude potential participants based on age, (actual or perceived) COPD-related health, or (actual or perceived) aptitude toward, or previous experience with, technology.

Most outcome measures tested were found to be suitable for use in a future RCT. The ISWT is routinely conducted in PR and would be suitable as an outcome measure for exercise capacity in a future trial. However, in this study, we were unable to identify a suitable outcome measure for physical activity. Fitbit One and Fitbit Charge were both accurate in counting participants’ steps during the F2 ISWT (when compared with Axivity sensor readings). However, control participants were more likely to have gaps in their step count data, and so these data are unlikely to have constituted an adequate comparison if this was an RCT. In addition, although control group participants could not see their step count on the Fitbit screen or on the smartphone’s home screen, some control group participants discovered the Fitbit app on the smartphone (a necessity for recording step count data) and were able to see their step counts. This meant that some control group participants were monitoring their step counts despite the blinded activity tracker. Thus, it would currently be challenging to use step count as a between-group comparative outcome measure in any future study. Participants also experienced difficulties completing the CHAMPS questionnaire for physical activity in older adults, thus ruling out this option as an outcome measure for physical activity. One physiotherapist suggested using a blinded activity tracker to take a baseline step count for participants between their assessment visit and beginning PR, which could provide a more reliable indication of physical activity changes within individual participants rather than between experimental groups in an RCT.

Some (though not all) participants were disappointed to be assigned to the control group, and in some cases, it was difficult for participants to appreciate the purpose of the control condition. For a future study, this may be resolved by giving all control group participants the opportunity to try using a wearable activity tracker at some stage in the study, eg, after data collection.

Implications for Future Development of the Intervention

Smartphones and wearables are two technologies predicted to transform health care provision in the coming decades [21]. Participants who helped codevelop the SMART-COPD
intervention stressed the importance of being able to personalize the intervention. This led to the inclusion of three different strategies to encourage maintenance of physical activity. However, when using the intervention in the real world, this approach proved too complicated. The daily walk and exercise components of the app were not widely used, and in some cases proved confusing for participants and PR team members alike. The more complex a technological intervention is, the less usable it is likely to be [26], and problems with usability may negatively impact participants’ motivation to continue with the intervention or the behavior [18]. Although efforts were made to design the intervention based on the needs and capabilities of people with COPD, more could have been done to achieve a truly co-designed intervention [55]. In addition, the self-selected participants who helped codevelop the SMART-COPD intervention were mostly experienced with digital technology and may not have been fully representative of the general COPD population.

Our results also suggest that people’s engagement with technology might wane over time (as had happened with some of our participants), although this study does not provide any indication as to whether participants began disengaging from the technology only or from the entire health behavior. This pattern was not present for all participants and the use of different components of the intervention differed between participants, which implies there were individual differences in how participants interacted with, and responded to, the intervention.

This study shows that participants also need the capability [18] to learn how to use the technology. People with less experience of digital technologies may have had less capability to use the intervention, and thus were more likely to leave the study. Participants appeared to have differing experiences of the SMART-COPD intervention depending on their previous experience with digital technology and their baseline health. This finding is indicative of the need to consider the circumstances, motivations, and capabilities of individual participants. The SMART-COPD intervention was complex, and the COPD population is complex and has complex needs. One intervention does not fit all, and future investigations of similar mHealth-based interventions need to consider individual factors as well as group factors when determining who could gain the most benefit [19].

Our results suggest that a future version of the app would need to be simplified. One option would be to adopt an existing commercially available activity tracker (eg, Fitbit) and its associated app to monitor step count and the completion of individually relevant step count goals. This approach has the benefit of not needing to be updated or maintained directly by the (resource-limited) research team, instead relying on a technology that is widely available and has a commercial team developing and maintaining it. However, while the SMART-COPD app was specifically designed to consider the needs of users with COPD and included simplification of the presentation of step count information collected via the activity tracker, commercially available apps are not designed for this population. Hence, we would need to determine whether patients with COPD without prior experience with digital technology could benefit from this approach (and, indeed, to ensure that the motivational aspects of commercial apps are not harmful or counter-productive for this population).

Strengths and Limitations

One strength of the feasibility study was that the SMART-COPD intervention was tested with people with COPD in a real-life setting, with real-life complexities and challenges, and over a period of several months. The study involved both people with COPD and relevant HCPs and tested out design elements of an RCT. In addition, to our knowledge, no one has previously used a wearable activity tracker to monitor step count for both an experimental and a control group.

Potential limitations of the study include technical issues affecting participants’ experiences of using the intervention, and the effect of those issues on data completeness. The study also experienced a high dropout rate. While this in itself provided valuable information on the usability and acceptability of the intervention, it also affected data completeness. We were unable to formally interview these participants to fully understand their experiences with the technology and their reasons for leaving the study.

The resource-limited nature of the feasibility study meant we were only able to include three PR sites: these sites may not have been representative of PR services (and COPD populations) across the United Kingdom. To practically conduct the study, we also had to use more than one researcher for recruitment and data collection. Researchers were trained on the use of the technology and on how to introduce both the study and the technology to participants. However, there may still have been differences in how each researcher introduced and explained the technology and the study, which may have influenced some participants’ interactions with the technology.

However, all of the above strengths and limitation issues reflect the wider complexities of assessing mHealth interventions in real-world settings.

Conclusions

Overall, the SMART-COPD intervention was well liked and perceived as easy to use and easy to incorporate into participants’ daily lives by those who completed the study. However, there was a high dropout rate which implies high rates of people who were eligible for the intervention but who did not easily adopt the technology, or else disliked the study design (eg, because of allocation). The data suggest that people with COPD who had worse baseline health were more likely to withdraw from the study, which may indicate that this patient group is harder to reach with mHealth interventions. The results suggest the intervention would need to be simplified for future use, eg, by focusing on step count only, with the possible sole use of a wearable activity tracker and an associated app. This finding contradicts a key finding from our earlier codevelopment work, which emphasized the importance of having a multi-option personalizable intervention. In a future RCT, the control group would be offered an opportunity to use the intervention and either the ISWT or a within-subject measure of step count should be considered as a primary outcome measure. Any future evaluation of the intervention would need
to consider individual factors that affect the usability, acceptability, and efficacy of the intervention.

Acknowledgments
This paper presents the independent research by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber. The views expressed are those of the authors, and not necessarily those of the NIHR or the Department of Health and Social Care.

The authors would like to acknowledge the contributions of partner organizations and particularly those of the participants from this study.

Authors' Contributions
MH was the Principal Investigator on the project. MH, CB, YB, GM, and JP were involved in the study design and development of the protocol. CB and LP were involved in the day to day management of the project. SP and JF were technical researchers on the project involved in development, maintenance, and delivery of the intervention. YB, GM, and JP were advisors on the project. LP, CO, JB, RC, HD, and CB were involved in participant recruitment and data collection. CB, SP, JP, and MH were involved in data analysis and interpretation. CB wrote and finalized the manuscript. LP, SP, JP, GM, YB, and MH all made substantial contributions to the final manuscript. All authors reviewed the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Chronic Obstructive Pulmonary Disease Patient Interview Guide.
[DOCX File, 29 KB - mhealth_v8i6e16203_app1.docx ]

Multimedia Appendix 2
Healthcare professional interview guide.
[DOCX File, 29 KB - mhealth_v8i6e16203_app2.docx ]

Multimedia Appendix 3
Screen Flow for SMART COPD App.
[PPTX File, 1652 KB - mhealth_v8i6e16203_app3.pptx ]

Multimedia Appendix 4
CONSORT-eHealth Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1564 KB - mhealth_v8i6e16203_app4.pdf ]

References

https://mhealth.jmir.org/2020/6/e16203


41. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]


Abbreviations

BCW: Behavior Change Wheel
CC: chronic condition
CHAMPS: Community Healthy Activities Model Program for Seniors
COPD: Chronic Obstructive Pulmonary Disease
HCP: health care professional
ISWT: Incremental Shuttle Walk Test
mHealth: mobile health
NHS: National Health Service
NIHR: National Institute for Health Research
PR: pulmonary rehabilitation
RCT: randomized controlled trial
SMART: Self-Management supported by Assistive, Rehabilitative, and Telehealth technologies
SUS: System Usability Scale
Accuracy of Distance Recordings in Eight Positioning-Enabled Sport Watches: Instrument Validation Study

Rahel Gilgen-Ammann¹, PhD; Theresa Schweizer¹, BSc; Thomas Wyss¹, PhD
Swiss Federal Institute of Sport Magglingen, Magglingen/Macolin, Switzerland

Corresponding Author:
Rahel Gilgen-Ammann, PhD
Swiss Federal Institute of Sport Magglingen
Hauptstrasse 247
Magglingen/Macolin,
Switzerland
Phone: 41 584676321
Email: rahel.gilgen@baspo.admin.ch

Abstract

Background: Elite athletes and recreational runners rely on the accuracy of global navigation satellite system (GNSS)–enabled sport watches to monitor and regulate training activities. However, there is a lack of scientific evidence regarding the accuracy of such sport watches.

Objective: The aim was to investigate the accuracy of the recorded distances obtained by eight commercially available sport watches by Apple, Coros, Garmin, Polar, and Suunto when assessed in different areas and at different speeds. Furthermore, potential parameters that affect the measurement quality were evaluated.

Methods: Altogether, 3 × 12 measurements in urban, forest, and track and field areas were obtained while walking, running, and cycling under various outdoor conditions.

Results: The selected reference distances ranged from 404.0 m to 4296.9 m. For all the measurement areas combined, the recorded systematic errors (±limits of agreements) ranged between 3.7 (±195.6) m and –101.0 (±231.3) m, and the mean absolute percentage errors ranged from 3.2% to 6.1%. Only the GNSS receivers from Polar showed overall errors <5%. Generally, the recorded distances were significantly underestimated (all P values <.04) and less accurate in the urban and forest areas, whereas they were overestimated but with good accuracy in 75% (6/8) of the sport watches in the track and field area. Furthermore, the data assessed during running showed significantly higher error rates in most devices compared with the walking and cycling activities.

Conclusions: The recorded distances might be underestimated by up to 9%. However, the use of all investigated sport watches can be recommended, especially for distance recordings in open areas.

Introduction

Background

There are many wearable devices on the market, especially in the health and sports sectors, that can access global navigation satellite system (GNSS) information [1]. A world survey of fitness trends identified wearable technologies such as GNSS-enabled watches and activity trackers as a key trend in 2016, 2017, and 2019 [2]. Conventional GNSS-enabled sport watches are predominant for a diverse population of active runners of different fitness levels [3,4]. For example, the wearable technologies used by runners during a half-marathon and marathon competition were as follows: 44.7% (437/977) were represented by GNSS-enabled sport watches and 18.5% (181/977) by mobile phones with a combined app to track running performance. In comparison, the proportions of heart rate monitors (37/977, 3.8%), wristband activity trackers (27/977, 2.8%), and smart watches (14/977, 1.4%) were quite low during these competitions. Wiesner et al [4] revealed that within runners using wearable technologies, the most frequent parameters of interest were the distance covered (523/617,
84.8%), time (441/617, 71.5%), and average speed (412/617, 66.8%). In that study, 3 out of 4 participants stated that they always trusted the data. As the users rely on these data to guide their training or competition, monitor their training volume, or plan their exercises, knowledge about GNSS accuracy is of importance [5].

Prior Work
In a systematic review of mobile apps to quantify aspects of physical activity, 20% (5/25) of the studies investigated the validity of mobile apps when measuring distance using GNSS information [6]. Mean percentage errors ranged between 2% and 10%. A systematic review of the validity of consumer-wearable activity trackers in 2015 revealed 22 studies [7]. However, only one study reported information on recorded distance but not using GNSS information. Recently, Pobiruchin and coworkers [3] investigated the recorded distance data obtained from different GNSS-enabled devices and brands during a half-marathon competition. They revealed small mean absolute errors of 0.12 km (0.6%) during the 21.1 km course. In the only study investigating GNSS-enabled sport watches, results from the validity of the recorded distances showed 0.8%, 1.2%, and 6.2% error rates on a straight path with open sky, an urban path, and a forest path, respectively [5]. However, further investigations are needed on recorded distances obtained in standardized settings to learn about different brands and products in the sport sector. Different sport watches should be investigated simultaneously in various real-world scenarios, both area-wise and speed-wise [8-10].

Difficulties in the Global Navigation Satellite System
To better understand why there might be difficulties in the accurate assessment of distance traveled by GNSS-enabled devices, one must comprehend how such devices work and what the GNSS signal affects, and therefore, how this impacts the measurement quality. Four main satellite implementations exist: GPS (United States), Global Navigation Satellite System (GLONASS, Russia), Galileo (European Union), and BeiDou (China). The number of satellites for GPS, GLONASS, Galileo, and BeiDou are 31, 27, 22, and 19, respectively, which circle the Earth twice a day in a precise orbit at an altitude of approximately 20,000 km [11-15]. Each satellite transmits a unique right-hand polarized signal and orbital parameters that allow GNSS-enabled devices to decode and compute the precise location of the satellite. The GNSS receiver measures the distance to each satellite by the amount of time it takes to receive a transmitted signal to exactly locate the user’s position on Earth [15]. Several factors affect the signal transmitted between the satellites and the GNSS receiver such as bad signal acquisition, number of satellites, signal multipath, satellite geometry, and GNSS receiver clock errors [1,15-17]. Bad signal acquisition can happen if the user of a GNSS-enabled device disregards any of the manufacturers’ main instructions to achieve a high GNSS signal: staying outside, regularly synchronizing the watch to the mobile app or computer to download the latest satellite data (= assisted GPS data), updating the watch’s GNSS setting for whatever activity, or choosing GPS + GLONASS or Galileo. The more satellites a GNSS receiver can detect, the better the accuracy. To calculate one’s 2D position (latitude and longitude), a GNSS receiver must be locked on to the signal of at least three satellites. Therefore, the user should remain stationary with the watch facing up during signal acquisition. Furthermore, the signal multipath and satellite geometry affect the transmitted signal. A user may get position errors or no position readings at all when a signal is blocked. This can occur because a GNSS signal does not penetrate any solid constructions or water. In addition, the GNSS signal is reduced by dense vegetation or cloudy weather or near objects and buildings, as there are reflections that transform the right-hand polarization into left-hand polarization before it reaches the GNSS receiver. Generally, the satellite signals are more effective when the satellites are located at wide angles relative to one another. Therefore, during signal acquisition, the user should stay away from large buildings and dense vegetation and, ideally, remain in a flat open area. Last, measurement quality can be hampered by timing errors the GNSS receiver might have because it is less accurate than the atomic clocks on GNSS satellites. The user, however, cannot change the clock errors in the GNSS receiver. In the northern hemisphere, the ground stations’ determined locations can vary due to the mentioned error sources. With GPS, GLONASS, and GPS + GLONASS, the determined horizontal (and vertical) location errors can be 8.0 (SD 17.1) m, 9.4 (SD 18.3) m, and 7.1 (SD 14.0) m, respectively, with a 95% confidence interval [18-20]. Overall, little to no information about positioning accuracy is provided by the most common manufacturers of GNSS-enabled sport watches. Scott et al [21] rated measures of validity of GPS in team sports as good (<5%), moderate (5% to 10%), or poor (>10%).

Aims of the Study
The aim was to investigate the accuracy of and parameters affecting the recorded distances obtained by eight sport watches from Apple, Coros, Garmin, Polar, and Suunto when assessed in different areas and at different speeds under different external circumstances.

Methods
Study Setting
In this instrument validation study, measurements were conducted in three areas while performing three different speed categories [10] (Figure 1):

- Urban area: in the city center of Biel/Bienne (Switzerland) at 434 m above sea level on a flat street with narrow and partly high buildings
- Forest area: in terrain in Magglingen (Switzerland) at 905 m above sea level on uphill (total gradient of 52 altitude meters and 11% slope), downhill, and flat paths with partly tall trees
- Track and field area: in an open track and field stadium in Magglingen without a tribune at 954 m above sea level on a 400 m track in the middle of lane 1 without any satellite visibility constraints
Figure 1. The urban (1), forest (2), and track and field (3) measurement areas. White circles divide the courses into subsections that were randomly combined and added up to result in different selected reference distances within the same setting.

The courses in each area were marked to allow ideal replications of the measurements. All courses in a respective area were split into subsections that were randomly combined within one measurement area resulting in selected reference distances of 400 m to 4500 m. This was chosen to ensure variance of traveled distances as a possible independent variable within the same controlled setting and investigate whether the distance per se and number of travel direction changes affected the results.

Subsections of each course were accomplished partly or entirely or repeated in either direction (including U-turns), or any combination of these (Figure 1). A trundle wheel [22] was used as the reference measure for all selected subsections. Each subsection was assessed twice with the trundle wheel with an accuracy to 1 cm.

Measurements were taken in the three speed categories—walking, running, and cycling—to represent low-,
moderate-, and high-gait speeds [1]. These three speed categories were self-paced by the subject but steady—according to the subjective feeling—within one measurement. The cycling was performed using an electric bike (e-bike) [23] to ensure high speeds and steady and straight riding, particularly on the uphill section in the forest area. Self-pacing and different subsections were chosen to ensure data acquisition that represented different real-life situations. The activity task itself and steady speed were secondary, as the primary aim was to validate recorded distances. Also, having a range of reference distances allowed statistical analyses with normally distributed data.

Table 1. Investigated sport watches and their specifications.

<table>
<thead>
<tr>
<th>Sport watch model</th>
<th>Abbreviation</th>
<th>Manufacturer</th>
<th>Serial number</th>
<th>Firmware or OS version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Watch Series 4</td>
<td>AW4</td>
<td>Apple Inc</td>
<td>FH7XD28MKDH9</td>
<td>5.2</td>
</tr>
<tr>
<td>Apex 46 mm</td>
<td>CoA</td>
<td>Coros Wearables Inc</td>
<td>73F855</td>
<td>1.31</td>
</tr>
<tr>
<td>Fenix 5X Plus</td>
<td>G5X+</td>
<td>Garmin Ltd</td>
<td>5MM005560</td>
<td>6.0 (7.6)a</td>
</tr>
<tr>
<td>Forerunner 935</td>
<td>GF935</td>
<td>Garmin Ltd</td>
<td>50S007800</td>
<td>12.50 (13.00)b</td>
</tr>
<tr>
<td>Vantage M</td>
<td>VM</td>
<td>Polar Electro Oy</td>
<td>464F832E</td>
<td>4.0.0</td>
</tr>
<tr>
<td>Vantage V</td>
<td>VV</td>
<td>Polar Electro Oy</td>
<td>4AF6F824</td>
<td>4.0.0</td>
</tr>
<tr>
<td>V800</td>
<td>V800</td>
<td>Polar Electro Oy</td>
<td>306AE719</td>
<td>1.11.49</td>
</tr>
<tr>
<td>Suunto 9 Baro</td>
<td>S9B</td>
<td>Suunto Oy</td>
<td>1.8251E+11</td>
<td>2.5.18 (2.6.54)a</td>
</tr>
</tbody>
</table>

aUpdates were needed during data collection to synchronize data to the respective website. Firmware version is listed after the update.

Subject
One healthy, fit, and lean person (female, aged 26 years, 53.0 kg, 1.58 m) performed all measurements. Having one subject ensured perfect standardized measurements. Moreover, independent variables such as body height, arm length and movement, walking pattern, etc, could be precluded. The subject was well familiar with the handling of all eight investigated watches and with the study design.

Data Collection
Measurements were scheduled for different days (7:00 am to 6:00 pm) from April to July 2019. On measurement days, all sport watches were synchronized to the respective mobile app and computer software by the supervisor to download the latest satellite data.

During each measurement, four watches were worn simultaneously. The subject wore two watches per forearm at least 4 cm apart to minimize potential interference between the devices (Figure 2). The combination of which four watches, the wearing side (wearing the watches on the right or left arm), and the wearing position (wearing the watch higher or lower on the forearm) were randomly assigned by the supervisor using a covariate adaptive randomization approach prior to the measurements [10,24]. It is worth noticing that the higher the watch is located on the arm, the stronger the signal-blocking effect due to the body’s interference. The watches were always mounted on bare skin and were not covered by sleeves. The subject prepared all the watches to receive the GNSS signal while standing in a flat, mostly open area without large buildings or dense vegetation. After readying each watch’s positioning connection, the subject waited another 5 minutes with arms outstretched, watches facing up, for calibration purposes to reach the best GNSS signal quality prior to starting the data acquisition (Figure 2).

Thereafter, the subject started the watches from left to right and did the same to stop the measurement. As the data collection was performed with one subject but eight watches, the same measurement—speed category and area with its combination of course subsections to reach the same reference distance lengths—was accomplished twice in a row, each time with four watches. Additionally, the following parameters were protocolled by the supervisor for each assessment: course turns per 1000 m, time of day, temperature, precipitation, sun, humidity, solar irradiance, and wind velocity [10]. In total, each of the three areas was completed four times in each of the three speed categories, resulting in $3 \times 4 \times 3 = 36$ measurements.

Instruments
Eight watches from the most common manufacturers in the field of sport watches (as of January 2019) were included in this study. The specific types were chosen based on personal communications with exercise physiologists and endurance athletes (see Table 1). All units were configured to the lowest possible 1-second (1 Hertz) GNSS recording, and the GPS + GLONASS satellite system was selected except for the Apple Watch Series 4 (Apple Inc), which does not have the option to choose the satellite system, and the V800 (Polar Electro Oy), which only has GPS due to its antenna implementation.
Figure 2. Randomly assigned wearing position on the forearm of the sport watches during one measurement (left). Calibration posture with arms outstretched to achieve the best global navigation satellite signal quality (right).

Data Processing
After each measurement, the data were uploaded to the respective software provided by and exported as default by the five investigated manufacturers. In the Garmin and Suunto devices, a firmware update was required during data collection to synchronize the data to the respective website. These data were not treated differently. To calculate the recorded distances of each watch, only the values of the real measurement time period were computed, except the data from the AW4, which could not be exported. In this case, the distance values shown on the AW4 display were noted, entered, and double checked in an Excel Windows 2016 (Microsoft Corporation) file by the supervisor for each measurement.

Statistical Analysis
Descriptive statistics with mean absolute and percentage errors, dependent samples 2-tailed \( t \) tests, Bland-Altman analyses, and a \( \pm 5\% \) accuracy of the recorded distance were used. The dependent \( t \) test was applied to test whether the difference of the recorded distances between tested devices and the reference values was zero. Bland-Altman analyses with corresponding 95\% limits of agreement (SD 1.96) were used to calculate systematic errors in the recorded distances. The \( \pm 5\% \) accuracy of the recorded distances was defined as the percentage at which the respective watch recording was within the proposed equivalence zone of \( \pm 5\% \) from the reference values [21,25]. Furthermore, multivariate linear regression analyses with stepwise backward elimination were used for each watch to detect independent variables with significant influence on the mean absolute error (MAE). The independent predictor variables investigated were speed categories, area, time of day, temperature, turns per 1000 m, precipitation, sun, humidity, solar irradiance, wind velocity, watch wearing side (0=right arm; 1=left arm), and watch-wearing position (0=higher; 1=lower on the forearm). These were chosen as potential predictors, as they occur during a user’s everyday life (where to move, at what pace, how curvy the terrain is, etc); the wearing position on the higher and lower forearm was included to demonstrate whether the setup in the study affected the result. The adjusted \( R^2 \) and \( \beta^2 \) were used to estimate the explained variances of the dependent variable by all the included variables and by each independent variable, respectively. In the case of multicollinearity (\( r \geq 0.80 \)) or the nonsignificant prediction of the MAE, the relevant variable was excluded from the regression analysis. Any \( P<0.05 \) was considered statistically significant, and the \( \alpha \) level was .05. The statistical analyses were applied using SPSS Statistics 25.0 (IBM Corporation) and Excel.

Results
Main Findings
In total, 100\% (36/36) of the measurements were recorded for each sport watch but in the S9B, 97\% (35/36) of the measurements were analyzed due to a technical failure during one assessment in the forest area. The walking, running, and biking was accomplished on average at 5.4 (SD 0.2), 10.2 (SD 0.7), and 17.6 (SD 2.6) km/h, respectively.

For all three measurement areas combined, the recorded systematic errors (limits of agreements) ranged between 3.7 (±195.6) m and −101.0 (±231.3) m for the V800 and CoA, respectively (Multimedia Appendix 1). The mean absolute percentage error (MAPE) ranged from 3.2\% to 6.1\% for the V800 and the S9B, respectively. Only the three GNSS receivers from Polar showed overall MAPEs <5\%. On average, the mean recorded distances within \( \pm 5\% \), when compared with the reference values, ranged from 80.6\% (29/36) in the V800 to 44.4\% (16/36) in the G5X+ (Figure 3). Overall, only the AW4 (\( P=.08 \)) and the V800 (\( P=.83 \)) showed no statistically significant differences from the reference distance.
Figure 3. Relative deviation of the distances recorded by the 8 watches compared with the reference distance. The red lines indicate the proposed equivalence zone (±5% of the mean); the boxplots’ lower and upper boundaries indicate the 25% and 75% quantiles of the distance data, respectively, and the middle notch indicates the median data value. The whiskers include all data points that fall within the 1.5 interquartile range of the 25% and 75% quantile values. Circles and stars indicate distance data points that lie beyond the 1.5 and 3 interquartile ranges, respectively.

Measurement Areas

Specifically, the recorded distances were significantly different from the measured distances in the forest (all $P<.04$), urban (all $P<.03$), and track and field areas (AW4, G5X+, and S9 all $P<.001$; Tables 2-4). The Bland-Altman analyses showed an underestimation by all watches in the forest and urban areas (except the overestimation in the V800) but an overestimation in the track and field area. Further, in all watches, the lowest MAE and good ±5% accuracy were recorded in the track and field measurements.

Table 2. Recorded distances and error rates of the eight sport watches obtained in the urban area when compared with the mean reference distance of 2046.4 (SD 1159.7) m (n=12).

<table>
<thead>
<tr>
<th>Watch</th>
<th>Recorded distance (m), mean (SD)</th>
<th>$P$ value</th>
<th>Systematic errors (m), (limits of agreement)</th>
<th>Mean absolute error (m), (mean absolute percentage error)</th>
<th>5% accuracy$^a$, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW4</td>
<td>1951.6 (1088.3)</td>
<td>.03</td>
<td>−94.8 (257.6)</td>
<td>108.4 (5.1)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>CoA</td>
<td>1899.5 (1085.6)</td>
<td>.001</td>
<td>−146.9 (237.4)</td>
<td>146.9 (7.5)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>G5X+</td>
<td>1939.9 (1114.9)</td>
<td>.003</td>
<td>−106.4 (186.4)</td>
<td>110.8 (5.9)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>GF935</td>
<td>1857.1 (1054.5)</td>
<td>.003</td>
<td>−189.3 (344.7)</td>
<td>189.3 (8.9)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>VM</td>
<td>1949.2 (1110.1)</td>
<td>.02</td>
<td>−97.9 (234.8)</td>
<td>109.8 (5.3)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>VV</td>
<td>1941.9 (1118.9)</td>
<td>0</td>
<td>−104.5 (171.6)</td>
<td>105.3 (5.4)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>V800</td>
<td>2134.6 (1227.1)</td>
<td>.003</td>
<td>88.2 (154.8)</td>
<td>89.4 (3.9)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>S9B</td>
<td>1868.7 (1006.9)</td>
<td>.02</td>
<td>−177.7 (438.7)</td>
<td>191.6 (8.5)</td>
<td>5 (42)</td>
</tr>
</tbody>
</table>

$^a$Percentage at which the distance recorded by each device was within 5% of the reference distance.
Table 3. Recorded distances and error rates of the eight sport watches obtained in the forest area when compared with the mean reference distance of 2111.6 (SD 1109.9) m (n=12).

<table>
<thead>
<tr>
<th>Watch</th>
<th>Recorded distance (m), mean (SD)</th>
<th>P value</th>
<th>Systematic errors (m), (limits of agreement)</th>
<th>Mean absolute error (m), (mean absolute percentage error)</th>
<th>5% accuracya, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW4</td>
<td>1969.8 (1031.9)</td>
<td>.008</td>
<td>–141.8 (297.8)</td>
<td>148.3 (6.9)</td>
<td>5 (41.7)</td>
</tr>
<tr>
<td>CoA</td>
<td>1944.1 (1032.6)</td>
<td>&lt;.001</td>
<td>–167.4 (177.2)</td>
<td>167.4 (8.5)</td>
<td>2 (16.7)</td>
</tr>
<tr>
<td>G5X+</td>
<td>1946.1 (1029.2)</td>
<td>&lt;.001</td>
<td>–165.4 (175.7)</td>
<td>165.4 (8.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>GF935</td>
<td>1983.0 (1037.5)</td>
<td>&lt;.001</td>
<td>–128.6 (171.3)</td>
<td>128.6 (6.0)</td>
<td>4 (33.3)</td>
</tr>
<tr>
<td>VM</td>
<td>1993.4 (1048.7)</td>
<td>&lt;.001</td>
<td>–118.2 (149.7)</td>
<td>118.1 (5.6)</td>
<td>6 (50.0)</td>
</tr>
<tr>
<td>VV</td>
<td>2000.4 (1047.4)</td>
<td>.001</td>
<td>–111.2 (157.8)</td>
<td>111.1 (5.0)</td>
<td>7 (58.3)</td>
</tr>
<tr>
<td>V800</td>
<td>2030.6 (1063.3)</td>
<td>.002</td>
<td>–81.0 (140.3)</td>
<td>81.0 (3.5)</td>
<td>9 (75.0)</td>
</tr>
<tr>
<td>S9B</td>
<td>1827.9 (988.1)</td>
<td>.07</td>
<td>–166.5 (529.4)</td>
<td>166.5 (7.5)</td>
<td>6 (54.5)b</td>
</tr>
</tbody>
</table>

aPercentage at which the distance recorded by each device was within 5% of the reference distance.
bn=11.

Table 4. Recorded distances and error rates of the eight sport watches obtained in the track and field area when compared with the reference distance of 2104.3 (SD 1167.4) m (n=12).

<table>
<thead>
<tr>
<th>Watch</th>
<th>Recorded distance (m), mean (SD)</th>
<th>P value</th>
<th>Systematic errors (m), (limits of agreement)</th>
<th>Mean absolute error (m), (mean absolute percentage error)</th>
<th>5% accuracya, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW4</td>
<td>2196.4 (1227.4)</td>
<td>.001</td>
<td>92.1 (137.6)</td>
<td>92.1 (4.1)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>CoA</td>
<td>2115.7 (1180.4)</td>
<td>.13</td>
<td>11.4 (47.3)</td>
<td>18.7 (0.9)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>G5X+</td>
<td>2165.4 (1190.1)</td>
<td>.001</td>
<td>61.1 (86.7)</td>
<td>61.1 (3.0)</td>
<td>10 (83)</td>
</tr>
<tr>
<td>GF935</td>
<td>2121.6 (1182.2)</td>
<td>.20</td>
<td>17.2 (86.3)</td>
<td>31.2 (1.3)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>VM</td>
<td>2142.3 (1214.5)</td>
<td>.049</td>
<td>38.0 (116.8)</td>
<td>48.7 (2.1)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>VV</td>
<td>2134.1 (1206.0)</td>
<td>.05</td>
<td>29.8 (92.7)</td>
<td>43.8 (2.3)</td>
<td>10 (83)</td>
</tr>
<tr>
<td>V800</td>
<td>2108.2 (1171.5)</td>
<td>.85</td>
<td>3.9 (134.6)</td>
<td>53.4 (2.3)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>S9B</td>
<td>2150.5 (1194.8)</td>
<td>.001</td>
<td>46.1 (65.4)</td>
<td>49.2 (2.5)</td>
<td>12 (100)</td>
</tr>
</tbody>
</table>

aPercentage at which the distance recorded by each device was within 5% of the reference distance.

Affecting Parameters

The backward multiple linear regression analyses on each watch revealed different significant predictors of an increased MAE (Table 5). The included independent variables explained between 18.3% of the variance in the MAE in the AW4 and 44.2% in the CoA. The running category was the most shown predictor; in six watches, it remained and had a significant influence on the final regression models and explained between <1% and 9% of the MAE in the respective watches.
Table 5. Linear regressions for each device separately with the mean absolute error as dependent variable.

<table>
<thead>
<tr>
<th>Watch and predictors</th>
<th>Interpretationa</th>
<th>Odds ratio (P value)</th>
<th>$R^2$</th>
<th>$F$ value</th>
<th>Explained variance, $\beta^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AW</td>
<td></td>
<td>.18</td>
<td>3.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>76.48 (.05)</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arm position</td>
<td>65.10 (.08)</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>CoA</td>
<td></td>
<td>.44</td>
<td>13.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Track and field</td>
<td>−138.47 (&lt;.001)</td>
<td></td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>59.49 (.05)</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>G5X+</td>
<td></td>
<td>.32</td>
<td>5.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Track and field</td>
<td>−104.34 (.002)</td>
<td></td>
<td>18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>49.65 (.07)</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forest</td>
<td>−54.60 (.08)</td>
<td></td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>GF935</td>
<td></td>
<td>.40</td>
<td>7.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>−14.11 (.002)</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>102.72 (.01)</td>
<td></td>
<td>16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time of day</td>
<td>394.14 (.02)</td>
<td></td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>VM</td>
<td></td>
<td>.21</td>
<td>4.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation</td>
<td>39.15 (.03)</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>62.92 (.03)</td>
<td></td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>VV</td>
<td></td>
<td>.43</td>
<td>4.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation</td>
<td>47.19 (.003)</td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>66.99 (.02)</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Solar irradiance</td>
<td>.11 (.05)</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Turns per 1000 m</td>
<td>5.73 (.05)</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cycle</td>
<td>48.94 (.07)</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V800</td>
<td></td>
<td>.39</td>
<td>6.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Precipitation</td>
<td>43.53 (.001)</td>
<td></td>
<td>28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cycle</td>
<td>49.59 (.03)</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Running</td>
<td>42.39 (.06)</td>
<td></td>
<td>&lt;1</td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td></td>
<td>.27</td>
<td>3.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>−24.25 (.004)</td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Humidity</td>
<td>−5.82 (.02)</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sun</td>
<td>−2.73 (.08)</td>
<td></td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

aThe mean absolute error was increased by the respective predictors.

Discussion

Principal Findings

The aim of this study was to evaluate the accuracy of the recorded distances of eight sport watches under different real-world environmental conditions for various speeds and reveal the predictors affecting measurement quality. Our results showed that the V800 was the most accurate watch overall, with a systematic error of 3.7 m, a MAPE of 3.2%, and 80.6% of all distance recordings within ±5% of the reference values. Notably, the V800 can use only GPS satellites due to its antenna implementation. Consequently, it is questionable whether the number of satellites and combination of different GPS + GLONASS or Galileo affects measurement quality that much, which was previously questioned [1]. In contrast, other devices showed a significant systematic error of up to −101.0 m and limits of agreements of over ±400 m, overall MAPEs of up to 6.1%, and less than 50% of the data falling within the tolerable range of ±5% (Figure 3). Overall, the recorded distances were underestimated in all watches, and the variance and some outliers were rather high. In contrast, during the Trollinger Half-Marathon, an overall MAPE of 0.6% was observed in the GNSS-enabled devices, of which the Garmin devices performed the most accurately [3]. In addition, the recorded distances...
generally showed an overestimation of the half-marathon distance. However, comparability with our study is limited, as our data were assessed under standardized conditions, whereas during the Trollinger Half-Marathon, all runners started/stopped and calibrated their devices individually and potentially did not run on the ideal route to complete the entire 21.1 km.

**Measurement Areas**

Considering the different measurement areas separately, an underestimation of the recorded distances in the forest and urban areas (except for the V800) was observed. In these areas, the MAPEs ranged from −3.5% to −8.9%, and a low 5% accuracy of 0% to 75% indicated large variances. This was in line with the previous research, demonstrating an underestimation of the recorded distances by −1.2% and −6.2% in an urban and forest area, respectively [5]. These results underline the fact that the GNSS signal is reduced in obstructed conditions such as dense vegetation or near objects and buildings, as it may reflect off before it reaches the GNSS receiver (ie, the GNSS not only receives signals directly from the satellites, signals are also reflected off such surfaces) [16]. In contrast, in the track and field area, the recorded distances were overestimated compared with the reference distance. However, the MAPEs were all <5% and ranged from 0.9% to 4.1% only. Furthermore, a good 5% accuracy was shown in the track and field area, with 5 devices having 100% (CoA, GF935, VM, V800, S9B), 2 devices having 83% (G5X+, VV), and 1 device having 67% (AW4) of the distance recordings falling within ±5% accuracy threshold. The authors assume that manufacturers may autocorrect the recorded distances in the first place to level out the underestimation in difficult areas, which may in turn result in an overestimation of the distance recordings in unobstructed conditions, such as flat and open areas [17].

**Affecting Parameters**

The included independent variables explained as much as from 18.3% to 44.2% of the variance in the MAEs of the distance recordings. Additionally, the running category showed in 75% (6/8) of the sport watches significantly increased error rates in recorded distances when compared with walking and cycling speeds. We assume this error is related to the gait-induced arm swing than to the movement speed itself. In comparison, cycling was the activity with the highest absolute speed, but in 25% (2/8) of the watches, it remained in the final regression model only. Previously, research compared the recorded altitude gains when assessed in the same brand on sport watches simultaneously placed on the wrist and on the hip when walking or running [26]. The watch placed on the hip was always more accurate than the watch placed on the wrist, and the error in altitude measures increased with faster movement speed. It was argued that the gait-related arm swing negatively affected the measurement accuracy which, in running, is raised in amplitude and frequency compared with walking. Furthermore, in our study, more rain was a significant predictor of an increased MAE in 38% (3/8) of the watches, which might be related to the impaired measurement accuracy in cloudy weather.

**Practical Implications**

Recent research highlighted the broad use of GNSS-enabled watches in runners of different fitness levels and that users trust the data of such devices [3,4]. However, our study showed, depending on what device was applied, that from 80.6% (29/36) to as little as 44.4% (16/36) of the mean recorded distances fell within ±5% when compared with the reference values. In particular, running over walking and cycling activities were shown to impair the GNSS accuracy in the recorded distances. Nevertheless, the use of all the investigated sport watches can be recommended, especially for distance recordings in an open area. Yet in case of training monitoring and regulation based on recorded distance data, one must be aware that the recorded distances might be underestimated by up to 9%. As such, correct execution of the manufacturers’ instructions is essential to get the best accuracy (ie, for the latest satellite data to be valid).

**Limitations**

Although we controlled for the wearing side and wearing position of the sport watches, we cannot exclude potential interference between the devices [10]. Only 14% (5/36) of the measurements were accomplished with moderate to heavy precipitation. In addition, the independent variables watch-wearing side and watch-wearing position occurred in a limited number of measurements only. Therefore, the power in the regression analysis is reduced, which in turn diminishes the interpretation of these predictors of increased error rates. Our data acquisition was performed by a single subject to ascertain perfect standardization. However, we cannot exclude that a study sample with different anthropometrics would fully affirm our results. Last, the selection of the specific eight sport watches might be biased as it was based on personal communications with exercise physiologists and endurance athletes rather than based on a detailed market research.

**Conclusions**

Our results showed that there was an overall moderate to good GNSS accuracy regarding recorded distances, with MAPEs ranging from 3.2% to 6.1% when assessed in urban, forest, and track and field areas. However, only three of the eight investigated GNSS-enabled sport watches reported an average MAPE <5%. Noticeably, in the unobstructed conditions of an open area, 75% (6/8) of the sport watches were able to accurately record distances, whereas in the obstructed conditions of forest and urban areas, this accuracy was limited, with a general underestimation of the covered distances. Furthermore, the data assessed during running showed significantly higher error rates in most devices compared with the walking and cycling activities.

**Acknowledgments**

Polar Electro Oy (Finland) funded this experiment in part. Polar Electro Oy provided the Swiss Federal Institute of Sport Magglingen (SFISIM) with financial support to conduct the study. The funding was targeted for data collection, results analysis, and Polar reporting costs. Additionally, the products tested were provided by Polar. The Polar products were from the company stock.
directly and the other products were bought by Polar from stores and given to us for the period of the study. After termination, all products were returned to Polar. As agreed beforehand, representatives from Polar Electro Oy had no influence on the data collection or analysis or on the outcome of the article or any right to stop the SFISM from publishing the findings. The manuscript content does not necessarily reflect the views of Polar Electro Oy.

**Authors’ Contributions**

RGA and TW conceived and designed the research. TS conducted the experiments. RGA analyzed the data and wrote the manuscript. All the authors read and approved the manuscript.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Bland-Altman plots for each device distinguished by measurement areas and speed categories.

[PNG File, 261 KB - mhealth_v8i6e17118_app1.png]

**References**


Abbreviations

AW4: Apple Watch Series 4 (Apple Inc)
CoA: Apex 46 mm (Coros Wearables Inc)
GSX+: Fenix 5X Plus (Garmin Ltd)
GF935: Forerunner 935 (Garmin Ltd)
GLONASS: Global Navigation Satellite System
GNSS: global navigation satellite system
MAE: mean absolute error
MAPE: mean absolute percentage error
S9B: Suunto 9 Baro (Suunto Oy)
V800: V800 (Polar Electro Oy)
VM: Vantage M (Polar Electro Oy)
VV: Vantage V (Polar Electro Oy)

©Rahel Gilgen-Ammann, Theresa Schweizer, Thomas Wyss. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 24.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Correction: Evaluating the Feasibility and Acceptability of a Mobile Health–Based Female Community Health Volunteer Program for Hypertension Control in Rural Nepal: Cross-Sectional Study

Zhao Ni¹, PhD; Namratha Atluri¹, BSc; Ryan J Shaw¹, PhD; Jingru Tan², MPH; Kinza Khan¹; Helena Merk¹, BSc; Yunfan Ge³, MPH; Shrinkhala Shrestha³, MPH; Abha Shrestha³, MD; Lavanya Vasudevan¹, PhD; Biraj Karmacharya³, PhD; Lijing L Yan², PhD, MPH

¹Duke University, Durham, NC, United States
²Duke Kunshan University, Kunshan, China
³Kathmandu University, Kathmandu, Nepal

Corresponding Author:
Biraj Karmacharya, PhD
Kathmandu University
BP Highway, Panauti 45209
Kathmandu
Nepal
Phone: 977 11 490497
Email: birajmk@gmail.com

Related Article:
Correction of: https://mhealth.jmir.org/2020/3/e15419/
doi:10.2196/19048

The authors of “Evaluating the Feasibility and Acceptability of a Mobile Health–Based Female Community Health Volunteer Program for Hypertension Control in Rural Nepal: Cross-Sectional Study” (JMIR Mhealth Uhealth 2020;8(3):e15419) noticed a typo in the Abstract of their published manuscript.

The following sentence:

The goal of this study was to assess if a mobile health–based female community health volunteer approach of combining the traditional community health volunteer program with digital technologies would be feasible and acceptable in rural Nepal.

Has been revised to:

The goal of this study was to assess if a mobile health–based female community health volunteer program with digital technologies would be feasible and acceptable in rural Nepal.

The correction will appear in the online version of the paper on the JMIR website on June 11, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
Corrigenda and Addenda

Correction: Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the “Needle in a Haystack”

Florian Dittrich1, Dr Med; Sascha Beck1,2, Dr Med; Anna Katharina Harren3, Dr Med; Felix Reinecke4; Sebastian Serong1, Dr Med; Jochen Jung5, Dr Med; David Alexander Back6, PD; Milan Wolf1; Stefan Landgraeber1, Prof Dr

1Department for Orthopaedics and Orthopaedics Surgery, Saarland University Medical Center and Saarland University Faculty of Medicine, Homburg, Germany
2Sportsclinic Hellersen, Lüdenscheid, Germany
3Department of Plastic, Reconstructive & Aesthetic Surgery, Specialized Clinic Hornheide, Münster, Germany
4Center for Orthopedics and Traumatology, University Medicine Essen, Essen, Germany
5Department for Orthopedics, Diakonie Hospital Bad Kreuznach, Bad Kreuznach, Germany
6Clinic of Traumatology and Orthopedics, Bundeswehr Hospital Berlin, Berlin, Germany

Corresponding Author:
Florian Dittrich, Dr Med
Department for Orthopaedics and Orthopaedics Surgery
Saarland University Medical Center and Saarland University Faculty of Medicine
Kirrbergerstr 100
Homburg
Germany
Phone: 49 68411624520
Email: florian.dittrich@uks.eu

Related Article:
Correction of: https://mhealth.jmir.org/2020/5/e17085/
doi:10.2196/21600

In “Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the “Needle in a Haystack” (JMIR Mhealth Uhealth 2020;8(5):e17085) the authors noted an error.

The paper was incorrectly listed as the article type “Review.” The correct article type is “Original Paper.”

The correction will appear in the online version of the paper on the JMIR Publications website on June 24, 2020, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

http://mhealth.jmir.org/2020/6/e21600/ (page number not for citation purposes)
is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Testing Suicide Risk Prediction Algorithms Using Phone Measurements With Patients in Acute Mental Health Settings: Feasibility Study

Alina Haines-Delmont1, BA, BSc, MSc, PhD; Gurdit Chahal2, BA, MSc; Ashley Jane Bruen3, BSc, MSc; Abbie Wall3, BSc, MSc; Christina Tara Khan4, MD, PhD; Ramesh Sadashiv2, BE, MSc; David Fearnley5, MBChB, FRCP, MSc, MBA

1Faculty of Health, Psychology and Social Care, Manchester Metropolitan University, Manchester, United Kingdom
2CLARA Labs, CLARA Analytics, Santa Clara, CA, United States
3University of Liverpool, Health Services Research, Liverpool, United Kingdom
4Stanford University School of Medicine, Stanford, CA, United States
5Mersey Care NHS Foundation Trust, Prescot, United Kingdom

Corresponding Author:
Alina Haines-Delmont, BA, BSc, MSc, PhD
Faculty of Health, Psychology and Social Care
Manchester Metropolitan University
Brooks Building, Nursing Department
53 Bonsall Street
Manchester, M15 6GX
United Kingdom
Phone: 44 01612472461 ext 2461
Email: a.haines@mmu.ac.uk

Abstract

Background: Digital phenotyping and machine learning are currently being used to augment or even replace traditional analytic procedures in many domains, including health care. Given the heavy reliance on smartphones and mobile devices around the world, this readily available source of data is an important and highly underutilized source that has the potential to improve mental health risk prediction and prevention and advance mental health globally.

Objective: This study aimed to apply machine learning in an acute mental health setting for suicide risk prediction. This study uses a nascent approach, adding to existing knowledge by using data collected through a smartphone in place of clinical data, which have typically been collected from health care records.

Methods: We created a smartphone app called Strength Within Me, which was linked to Fitbit, Apple Health kit, and Facebook, to collect salient clinical information such as sleep behavior and mood, step frequency and count, and engagement patterns with the phone from a cohort of inpatients with acute mental health (n=66). In addition, clinical research interviews were used to assess mood, sleep, and suicide risk. Multiple machine learning algorithms were tested to determine the best fit.

Results: K-nearest neighbors (KNN; k=2) with uniform weighting and the Euclidean distance metric emerged as the most promising algorithm, with 68% mean accuracy (averaged over 10,000 simulations of splitting the training and testing data via 10-fold cross-validation) and an average area under the curve of 0.65. We applied a combined $5 \times 2 F$ test to test the model performance of KNN against the baseline classifier that guesses training majority, random forest, support vector machine and logistic regression, and achieved $F$ statistics of 10.7 ($P=.009$) and 17.6 ($P=.003$) for training majority and random forest, respectively, rejecting the null of performance being the same. Therefore, we have taken the first steps in prototyping a system that could continuously and accurately assess the risk of suicide via mobile devices.

Conclusions: Predicting for suicidality is an underaddressed area of research to which this paper makes a useful contribution. This is part of the first generation of studies to suggest that it is feasible to utilize smartphone-generated user input and passive sensor data to generate a risk algorithm among inpatients at suicide risk. The model reveals fair concordance between phone-derived and research-generated clinical data, and with iterative development, it has the potential for accurate discriminant risk prediction. However, although full automation and independence of clinical judgment or input would be a worthy development for those
individuals who are less likely to access specialist mental health services, and for providing a timely response in a crisis situation, the ethical and legal implications of such advances in the field of psychiatry need to be acknowledged.

(JMIR Mhealth Uhealth 2020;8(6):e15901) doi:10.2196/15901

KEYWORDS

suicide; suicidal ideation; smartphone; cell phone; machine learning; nearest neighbor algorithm; digital phenotyping

Introduction

Background

Limitations in scalability, accuracy, and consistency with respect to traditional methods of predicting suicidal behavior have been recognized in the literature and meta-analyses [1-5]. Suicidality has been defined as any suicide-related behavior, including completing or attempting suicide (intent), suicidal ideation (thoughts), or communications [6]. Not everyone who experiences suicidal ideation attempts suicide, but suicidal thoughts have been shown to be linked to a higher risk of death by suicide [7]. Although some people communicate their suicidal thoughts or plans to family and friends before suicide, others do not disclose their intent [8-10]. In addition, some individuals might not seek help during a time of crisis because of various perceived constraints, including fear of stigma or disclosure, lack of time, access to services, and preference for informal help [11]. Our ability to predict suicide is limited by our understanding of suicidal thoughts and their nature [12].

Suicidal Ideation, Smartphone apps, and Machine Learning

Advances in smartphones and connected sensors (wearables) have opened new possibilities for real-time, context-related monitoring of suicidal thoughts and suicidal risk [13], for example, ecological momentary assessments [14] that allow self-reporting of suicidal thoughts as they occur in an individual’s day-to-day life, naturalistic setting [15] and digital phenotyping that enables access to real-time classification and quantification of human behavior [16-18].

The use of computational data-driven methodologies that use social media to understand health-related issues (infodemiology, infouesveillance [19,20]) and data mining techniques (artificial intelligence, machine learning algorithms [21]) provides additional potential in expanding our understanding of people’s thoughts, feelings, behavior, etc and improving monitoring of suicide risk in real time. Although in its infancy, new research exploring suicidal ideation has shown that social media (eg, Twitter and Facebook) could potentially be used as a suicide prevention tool [10,22-26]. One study, for example, demonstrated the utility of social media blog post analysis in classifying individuals with high suicide risk in China [27]. Some research indicates that by analyzing certain patterns of smartphone use, changes in mental health symptoms could be identified [28].

Although standardized clinical tools can help to classify factors that contribute to suicide risk and understand biological markers related to suicide (trait analyses), computer science and machine learning can provide additional and timely tools to understand linguistic markers of suicide thought (state analyses) [29]. New statistical methods have been proposed and tested to achieve more accurate predictions of risk, for example, support vector machines (SVMs), deep neural nets, and random forests [13]. Evidence suggests that these methods, especially elastic net, perform better than traditional logistic regression techniques [30]. There is a shift toward developing more personalized risk profiles and using decision tree techniques that explore hundreds of predictors rather than a few clinically relevant risk factors [31]. Modern machine learning techniques are better placed to identify complex relationships between large datasets and suicide risk [13].

Early evidence generated by a pilot study using data from 144 patients with mood disorders suggests that machine learning algorithms using previous clinical data were successful in distinguishing between people that attempt suicide and those who do not, with a prediction accuracy between 65% and 72% [32].

Although there has been a growing body of research seeking to augment or advance traditional methods with the aid of machine learning in clinical psychiatry [2,4,29,30,33-37], the majority of studies rely on applying algorithms that learn from clinical data such as health care and electronic medical records, unstructured notes by providers and caretakers, or some other data carefully gathered by health care professionals.

Objectives

In this feasibility study, we aim to add to existing knowledge by using a nascent approach combining clinical data with proxy risk active and passive data collected from mobile devices to develop our algorithm. We developed a software platform to collect data on inpatients in acute mental health settings via our own mobile app, Strength Within Me (SWiM); a smartphone (iPhone); a wrist wearable (Fitbit; Fitbit, Inc); and questionnaires administered by the research team. Active risk data—patient-facing user interface modules (eg, journaling, safety plan, and mood meter)—and passive risk data that did not require direct interaction from the patient (eg, sleep monitoring) were collected behind the scenes. This information was then used to construct and train machine learning algorithms seeking to produce a risk score that deduces the likelihood of suicide. We used the risk level from the Columbia-suicide severity rating scale (C-SSRS) [38], which was assessed by mental health researchers as our standard classification target. C-SSRS is currently considered the gold standard approach for the measurement of suicidal ideation and behavior in clinical trials [39]. Previous research has confirmed the validity of the scale and its prediction accuracy for short-term risk of suicidal behavior in clinical and research settings. Studies have demonstrated that individuals who meet the criteria of high risk following the administration of C-SSRS are almost 4 times more
likely to attempt suicide within 24 months [38]. The C-SSRS was then compared with data from proxies for risk factors [40] such as sleep quality and emotional health collected via Fitbit () trackers and the SWiM app that patients interacted with for a week during their admission.

**Methods**

**Participants and Clinical Setting**

In this phase 1 feasibility study, we collected data from service users admitted to 6 acute adult mental health wards within a National Health Service trust in the North West of England, United Kingdom. Service users who had been admitted to a ward within the last 7-10 days were assessed by nursing staff to determine study eligibility. Following informed consent, participants were given a study iPhone and Fitbit to enable use of the SWiM app and monitor their sleep and daily activity for up to 7 days. Participants were then involved in 3 interviews at three different time points (ie, as soon as possible following admission, 3 days later and 7 days later or at discharge, whatever came first) to complete a battery of assessments, including the C-SSRS [38], examining suicidal thoughts and behavior. The interviews were completed by 2 experienced researchers who were trained to administer the clinical assessments. If suicidal risk was highlighted during the interview, nursing staff were informed and an agreed protocol was followed to ensure safety. Participants were given vouchers following the completion of assessments. In total, 80 patients out of the 186 eligible consented to participate and 66 were included in the analysis based on the completion of at least two follow-up clinical assessments. This represents a 43% response rate and 83% completion rate. For a breakdown of participants, see Figure 1.

**Figure 1.** Strength Within Me study flow diagram. Timeframe for recruitment: January-November 2018. Included in the analysis: participants who completed C-SSRS at second follow-up. C-SSRS: Columbia Suicide Severity Rating Scale.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects, ie, patients were approved by the Health Research Authority in England and the North West–Liverpool East Research Ethics Committee (Reference 17/NW/0173). Written informed consent was obtained from all inpatients.

**Overview of Participant Data Fed Into the Modeling Process**

At a high level, the data are segmented into (1) data entered by the user into the SWiM app; (2) data collected passively by the SWiM app, the Fitbit wearable, and the Apple Health app; (3) data directly gathered by our researchers; and (4) social interaction data for those who gave permission. We gathered a total of 173 variables—a mix of raw data such as counts of the number of journal entries and derived values that involve summary statistics or other variations of the data (eg, adding
up minute-wise sleep records to get total sleep time or the number of interruptions in sleep). Social interaction data were excluded from the analysis because of the low response rate, that is, 8 out of the 80 participants gave permission or had access to Facebook.

As shown in Table 1, user-inputted data included participant mood, free-form journal entries, steps for personal safety plans, and custom reminders they could have set for themselves. From these entries, we collected descriptive statistics such as average mood reported (Likert Scale 1-5), average character limit, maxima, minima, and raw counts. A particular derived variable of interest from journaling was the average sentiment derived for each journal entry. This sentiment (ranging from −1 for negative to +1 for positive) was calculated via a third-party model known as the Valence Aware Dictionary and Sentiment Reasoner [41], which is catered to sentiments expressed in social media but has proven itself in other domains. The idea behind using this model was to obtain a proxy for the indication of feelings by users as they write and reflect.

Data collected by the research team included sociodemographic information, such as age and gender, and clinical assessment data. The key information that we used in the modeling was the researchers’ assessment of the patient through the C-SSRS, which was assessed a maximum of 3 times (patient entry, 1 or two follow-ups during their hospital stay, and exit). All 80 consenting users were at risk upon entry to the ward (when the first test was done), so at this point, no prediction was done. The initial thought was to compare results against an intermediate survey result and exit survey result, considering the change in risk, but we did not have enough exit surveys for 2 different time period comparisons. Overall, 66 out of the 80 participants had taken at least a second survey where risk level was reassessed, and that was the population included for prediction. There was a 3-7 day wait from the first assessment to the second assessment.

Finally, we included passive data gathered via the phone and the Fitbit wearable, such as details about a user’s step frequency and count from Apple’s Health kit app, minute-level sleep data from Fitbit, and engagement patterns with the phone (eg, number of log-ins to the SWiM and the number of times a certain section was visited). Levels of engagement with study data are presented in Table 2.

### Table 1. Strength Within Me study data.

<table>
<thead>
<tr>
<th>Data source</th>
<th>Examples of variables collected</th>
<th>Examples of raw data</th>
<th>Examples of derived data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Stats of Facebook activity and post activity</td>
<td>Number of posts: 5 and number of total likes: 100</td>
<td>Average likes per post: 20</td>
</tr>
<tr>
<td>User input</td>
<td>Journal, mood, reminders, and safety plan steps</td>
<td>Journal entry: “Last night was horrible. I couldn’t sleep at all with the noise.”</td>
<td>Sentiment: −0.8 and word count: 12</td>
</tr>
<tr>
<td>Clinical team</td>
<td>Demographics and C-SSRS responses</td>
<td>Age: 35 years and C-SSRS risk overall: moderate</td>
<td>C-SSRS risk binary: 1</td>
</tr>
<tr>
<td>Passive sensor data</td>
<td>Sleep, steps, and interactions</td>
<td>[dateTime: 23:10:00, value: awake], [dateTime: 23:11:00, value: asleep]</td>
<td>Sleep latency: 1 min and average time asleep: 5 hours</td>
</tr>
</tbody>
</table>

C-SSRS: Columbia-suicide severity rating scale.

### Table 2. Engagement rate across active and passive data in the study (N=66).

<table>
<thead>
<tr>
<th>Data source</th>
<th>Rate, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step-related features (Fitbit and iPhone)</td>
<td>26 (40)</td>
</tr>
<tr>
<td>Journal entries (self-documented via SWiM app)</td>
<td>45 (68)</td>
</tr>
<tr>
<td>Mood entries (self-reported via SWiM app)</td>
<td>53 (80)</td>
</tr>
<tr>
<td>Phone activity (data usage)</td>
<td>66 (100)</td>
</tr>
<tr>
<td>Sleep (Fitbit)</td>
<td>59 (90)</td>
</tr>
</tbody>
</table>

SWiM: Strength Within Me.

### Modeling

**Machine Learning Setup and Data Analysis in Our Clinical Setting**

As a first step toward developing an algorithmic risk score that is valid in predicting suicide risk, we framed the problem as a supervised, a binary classification problem in which users were categorized in terms of levels of risk of low risk versus high risk using the information specified above. These low risk and high risk labels were derived from the overall C-SSRS risk scores obtained after asking participants a range of questions on previous attempts, ideation, etc. Usually the 3 categories are low, moderate, and high, but we grouped moderate and high for the sake of tractability from a modeling perspective. From a machine learning perspective, this aids in what is commonly referred to as the class imbalance problem [42], where certain categories have relatively few labels to their other counterparts. This makes it statistically more difficult to identify, and these
categories as models are inclined to achieve high scores by predicting the most common class; we turned a distribution of 36 low, 5 moderate, and 25 high to 36 low and 30 high. Choosing a binary case was helpful in dealing with the class imbalance issue, as models are data-dependent in terms of volume (i.e., the more examples, the better job they do in learning). This is especially critical when we take into account the limitations in our data; to fairly judge the model performance, we must partition the data (a test set and a training set via k-fold cross-validation [43]) to assess how well the model can predict risk on new users given what it is learned from old users [44]. From a risk-app perspective, although it would be ideal to place users on a continuum of risk levels, it is critical to first assess the feasibility of identifying users at discrete thresholds as well as seeing the degree to which we can match the current standard in risk assessment.

Our low risk and high risk categories were mapped to binary outputs of 0 or 1. Some features derived from a user’s journal entry are the word length and sentiment score (ranging from negative with −1 to positive +1; for further information, please refer to the source model from which this is derived [41]). To account for the time dependency in the data (multiple journal entries across multiple days for example), a majority of the features engineered were done so in a summary statistics fashion (mean, median, variance, etc). For example, the average journal word count per day over the user’s total number of entries was used to summarize one aspect of a user’s journaling behavior over their time with the app.

We curated 172 features formulated from categories of sleep data, journal entries, data usage, mood, and app activity statistics. For more information, a comma-separated values file including the full list of features incorporated into modeling (besides uid, which is user id to anonymize yet identify patient) is included in Multimedia Appendix 1. The 172 features were projected down to a 5-dimensional space by principal component analysis (PCA). This sample provides an insight into replicability. Any feature that has a summary statistic attached such as mean or std was done over the course of the 3-5 days before the second assessment. Categorical features such as gender were mapped to numerical (in this case binary) outputs for the algorithm to consume.

This is typically considered a relatively high number of features relative to the amount of possible supporting data points per number of users recorded. To provide a more suitable set from which a machine learning algorithm may distinguish a signal for risk, we turned to feature selection and dimensionality reduction techniques. Our aim was to cut down to a smaller set of features that may also be interpretable and grounded in clinical knowledge of risk factors. We, therefore, opted for PCA [45] as our dimension reduction technique and used Random Forest [46-48] to help in terms of feature selection as well to check the reliability of our reduction. Algorithms such as SVMs [49] are designed in such a manner as to overcome dimensionality issues, but they were experimentally confirmed to be unsuited to the task due to the size of the data.

For our study, the random forest model was composed of 25 decision trees. We took a look at the top 30 of the 170 original features and found that journal-related features such as average feeling, cell activity such as the variation in user’s data usage, sleep-related features such as average sleep efficiency (time spent sleeping or total time spent in bed), and other natural indicators, mostly known to clinical psychology, are markers of risk. For an example of a decision tree formed for our data, see Figure 2. The tree is read similarly to a flow chart in a top-down, left-right fashion. For example, at the top, we start with an entropy of 0.997 (entropy of 1 means complete uncertainty with 0 as certainty) [19-21], as we have 25 people in the low-risk category and 22 in the high-risk category. We, then, look at their average journal feeling, and if it is less than 3.161, we go to the left node with a subgroup of 34 people, otherwise the right node with a subgroup of 13 people. Following the right node, we now have a subgroup of 13 people with an average journal feeling greater than 3.161. On the basis of this characteristic alone, we reduce entropy to 0.619 (we are more certain of our group) and have 11 users correctly identified as high-risk, but 2 low-risk users misclassified as high-risk users. Reduce misclassification: we again split by the average amount of time the user has spent in bed. If they have greater than 541 min spent in bed in a day, a subgroup of 9 out of 9 people is correctly identified as high-risk users. However, we see that for less than 9 hours, we also predict high risk and have complete uncertainty (entropy 1), as the subgroup of 4 people is evenly divided among the classes. Once we reach 1 of these leaves or terminal nodes, we can read the decision process used to get there. For example, for the right-most leaf with 9 samples we discussed, users with an average journal feeling greater than 3.161, who also spend more than 9 hours in bed, are identified as high-risk users. Similar interpretations can be made for the other 6 terminal nodes. Worth noting is that the features are ordered top-bottom in terms of ability to split classes and reduce entropy; by this criterion, we see journal feeling as the most important feature, time in bed as second, and so on.

https://mhealth.jmir.org/2020/6/e15901

JMI Ro 8 e15901
Principal Component Analysis

For dimensionality reduction and to guard against overfitting, we turned to PCA. On a high level, PCA groups features that are correlated with one another into new features (principal components) that hold the most signal in terms of variation in the data [44]. The idea is that features that explain a high level of variability found in the data produce most of the signal needed to distinguish categories. Those features which do not contribute as too strongly are discarded; by doing so, we can obtain a concise set of features at the cost of a small drop in prediction performance. Formally, PCA is an orthogonal linear transformation that maps the data to a new coordinate system such that the bulk of the variance of the projection is covered by the first k components, where k < total number of original features and components are linear combinations of the originals. Another important characteristic of PCA is that it is not optimized for class separability and may be considered as an unsupervised model. This is critical as we aim to achieve generalizations outside of the data at hand, and we do not want to overfit our final model. To provide a visualization of the PCA transformation on our data, an example of a 2-dimensional, 2 components PCA is given in Figure 3.

After looking at the variance captured up to 100 possible components, we settled for the first 5 components, as they accounted for 55% of the variance. Our first 5 components are described in Table 3, along with the themes/patterns identified after reviewing which features were grouped. We were assured that these components made sense in terms of clinical knowledge of how sleep quality, mood, activeness, and other characteristics are indicators of mental health [50-53]. Moreover, the top 30 features of our feature selection from random forest strongly overlapped with these features, and so we were further assured in terms of potential predictive power.
Figure 3. A diagram of principal component analysis. A high-dimensional dataset has been flattened to a 2-dimensional space where the new axes correspond to the principal components (they point in the direction of the largest variance of the data).

Table 3. Principal components analysis components and patterns.

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
<th>Themes, patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>First component</td>
<td>Maximum efficiency, average efficiency, median efficiency, max time in bed, and number of sleep recordings</td>
<td>Ability to sleep, sleep quality</td>
</tr>
<tr>
<td>Second component</td>
<td>Number of packets sent, number of times connected to Wi-Fi, number of times connected to cellular data plan, and number of times journal entered</td>
<td>User app activity, data presence</td>
</tr>
<tr>
<td>Third component</td>
<td>SD sleep start, median journal feeling, max sleep start, max journal feeling, minutes in bed, and minimum journal feeling</td>
<td>Feeling versus sleep activity</td>
</tr>
<tr>
<td>Fourth component</td>
<td>Median char length, median word length, median journal feeling, SD rest duration, and max rest duration</td>
<td>Journal input versus resting variability</td>
</tr>
<tr>
<td>Fifth component</td>
<td>Median sentiment, SD number of awakenings during sleep, number of awakenings during sleep, and minimum sentiment</td>
<td>Sleep quality and reflection tone</td>
</tr>
</tbody>
</table>

Results

Prediction Algorithm Testing

We tested a series of algorithms that we thought would be best suited to predicting levels of risk from a theoretical perspective. Often referred to as the bias-variance trade-off [44, 54, 55], there is often the case with model selection that the best model should not be too simplistic such that its crude predictions miss a bulk of the cases, nor should it be overly complex such that its high sensitivity perfectly fits the data, but fails to generalize to new, unseen data. This principle, along with other individual algorithm properties, helped guide the experimentation. As discussed in the literature [56], increasing the complexity and flexibility of a model tends to allow it to understand more nuanced relations but at the cost of being overly sensitive to noise within data and overfitting. Hence, not only were models of varying complexities chosen for comparison from linear models such as logistic regression to nonparametric models such as K-nearest neighbors (KNN), but the parameters within each model were also tuned by choosing the number of neighbors and reducing dimensionality through PCA.

It is important to mention that these models are selected and judged based on various metrics that aim to capture the objective for which the model is needed. Certain metrics also have advantages over others depending on the imbalance of classes, nature of the data (categorical or numerical), and other factors. As we had a nearly balanced dataset, and this was a feasibility study, we opted for the simplest way to measure performance, in this case accuracy, to understand metric of accuracy where we measured the number of correctly predicted observations over the total number. As a baseline, we looked at the simplest heuristic of predicting the majority class of low-risk users. This produced an average accuracy of 53%. 
Random forest was tested as it is generally agreed upon as a strong out-of-the-box model that performs well on various datasets in different contexts as having interpretability through the feature importance it can help provide [46,48]. Logistic regression was another model considered due to the log-odds interpretability for the coefficients to each of the features (usually referred to as explanatory variables in explanatory contexts) and natural fit to classification problems [44]. SVMs [44,49] were also tested as they have the design of naturally combating the curse of dimensionality through the transformations they do to the data (kernel trick). SVMs are also rather sophisticated models that tend to produce near state-of-the-art results (barring neural networks which at the time of writing are highly data-hungry, and not necessarily interpretable). Finally, we considered the KNNs algorithm, which is often sought due to simplicity as well as the natural heuristic of classifying based on how close observations are to one another [57].

To test the performance, we performed k-fold cross-validation with k=10. This means that we randomly partitioned the data into 10 pieces (folds) and used 9 of them to train the model and 1 as an unseen piece (fold) to test on. This was done such that each of the 10 folds was used as the unseen/testing data at a given iteration. The idea was to obtain the expected accuracy of a model when exposed to new data by simulating variations of data seen to unseen data. We repeated this process 10,000 times to obtain a more stable estimate, as there are many ways to partition this data into 10 folds. Table 4 summarizes the results.

Table 4. The average cross-validation accuracy, along with the SD of the accuracy observed for the various folds.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>10-fold CV$^a$ average accuracy (10,000 iterations)</th>
<th>SD</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-nearest neighbors (k=2)+PCA$^b$ (n=5)</td>
<td>0.68</td>
<td>0.12</td>
<td>Best performance, k=2 seemed natural and worked best up to 10</td>
</tr>
<tr>
<td>Random forest (k=25)+PCA (n=5)</td>
<td>0.60</td>
<td>0.13</td>
<td>Nonlinear helps, too many trees did not, PCA reduced deviation</td>
</tr>
<tr>
<td>Random forest on raw features (k=25)</td>
<td>0.60</td>
<td>0.15</td>
<td>Nonlinear helps, too many trees did not</td>
</tr>
<tr>
<td>SVM$^c$ (degree 2 polynomial kernel)</td>
<td>0.57</td>
<td>0.10</td>
<td>Likely overfit, base guessing</td>
</tr>
<tr>
<td>Logistic regression+PCA (n=5)</td>
<td>0.59</td>
<td>0.14</td>
<td>Removed correlation due to PCA+prevent overfitting</td>
</tr>
<tr>
<td>Logistic regression on raw features</td>
<td>0.55</td>
<td>0.16</td>
<td>Likely overfit, base guessing</td>
</tr>
<tr>
<td>Baseline: guessing majority from training fold</td>
<td>0.53</td>
<td>0.20</td>
<td>Baseline to beat</td>
</tr>
</tbody>
</table>

$^a$CV: cross-validation.  
$^b$PCA: principal component analysis.  
$^c$SVM: support vector machine.

Logistic regression failed to perform much better than baseline. With the raw features, it performed poorly likely due to overfitting and high collinearity between some features (eg, median sleep time and mean sleep time). We removed most of this through PCA and performed slightly better on average at 59%, but the SD of 14% was worrying, given its below baseline lower end (worse than majority guessing). Similarly, SVM failed to perform much better, and of the different kernels, we present the polynomial degree 2 kernel as it performs best out of other variations (higher order polynomials, radial basis kernels, and linear). We defer the explanations of these kernels to the literature. Random forest performed better than either of the other 2 algorithms, but worst-case folds still fell below baseline.

The k-Nearest Neighbors Algorithm

The most promising was the KNNs algorithm with k=2 with the 5 principal components discussed earlier as features with not only an average accuracy of 68% (averaged over 10,000 simulations of splitting the training and testing data via 10-fold cross-validation) but with an SD of 12%, which resulted in its worst performance just above baseline at 56% and upper limit of 82% (Figure 4). In terms of false positive and true positive rates, the model achieved an average AUC of 0.65 (Figure 5).

We applied a 5×2 cv combined F test to test the model performance of KNN against the baseline classifier that guesses training majority, random forest, and other models such as supported vector machine and logistic regression and achieved F statistics of 10.7 (P=.009) and 17.6 (P=.003) for training majority and random forest, respectively, rejecting the null performance being the same [58,59]. Due to the promising results of the algorithm, we explain it to readers unfamiliar with it. The KNN algorithm essentially follows the saying of birds of the same feather flocking together. That is to say, the way prediction is performed using this algorithm is that for a new test point, the distance (usually the well-known Euclidean distance) is computed between the new point and k of the closest previously labeled observations. Of the k neighbors, the majority class is chosen as the label for the test point. For example, with k=5, we look at the features of a new person whose risk has not been identified yet and look at 5 people with the features closest matching this new person out of the training set. If 3 of them are high risk and 2 are low risk, the new person is identified as high risk, with 3 votes to 2. For even numbers of k, such as 6, where there might be ties, we weight the votes by proximity. Therefore, with respect to our PCA features, we are comparing people who have similar sleep characteristics, data usage, and so on. We found k=2 to perform best in our scenario, likely due
to the low sample size as well as high variability among users. We used a Euclidean distance metric and enforced each feature to have equal distance weighting (uniform weights).

Figure 4. Example of nearest neighbors with k=2 with data in 2 dimensions. Here, the new test point is x and has 1 minus neighbor and 1 plus neighbor as its 2 closest neighbors. As the minus neighbor is closer, the new point x will be classified as minus. “+” stands for positive class, “-” for negative class, and “x” for new data point that has yet to be assigned a class.

Figure 5. Receiver operating characteristic curve for k-nearest neighbors. AUC: area under the curve.

Discussion

Principal Findings

The results from this feasibility study indicate that, although not a perfect predictor, the KNN model is suitable for this study because it has shown the ability to separate users deemed at risk of suicide from the C-SSRS to those not deemed at risk at an average rate beyond just randomly guessing (ie, at an average rate 15% beyond randomly guessing the majority to be at low risk). These are early indications that it is possible to predict risk using the data collected in this feasibility study, using the KNN algorithm. The data used to inform this included users’ sleeping activity, step activity, self-reported mood, journaling thoughts, and activity levels as measured by a phone app.

This is a crucial first step in automatic risk assessment, as we managed to build an algorithm that predicts suicidal risk at a rate significantly better than the baseline of guessing the simple majority that were collected directly from smartphone interactions. This is also promising, as we are working with a relatively small dataset from a machine learning perspective. This is the basis for future phases of this study, where we will be looking to test the model on additional users of mental health services for further testing of concept and generalizability.

The implications of this feasibility study are highly significant for building capacity for suicide risk prediction (future risk) or detection (real-time, current risk). With a low proportion of suicide attempters who actually access mental health services [60], it is essential to develop and test nonclinical means of assessing risk. Given the dynamic nature of suicide ideation and suicide risk, new methods are needed to track suicide risk in real time [61], together with a better understanding of the ways in which people communicate or express their suicidality...
Mobile apps could be better suited to help prevent suicide by offering support in situ and at the time of crisis [62]. Although previous studies have utilized electronic health care record data to create an actuarial model of suicide risk [30,34,35,37,63-65] or focused on a single aspect of user input such as language [29,30,36], this study adds to the literature by introducing external, user-generated input, and smartphone data and combining it with clinical data. Our study adds to evidence that reports on the use of external, nonclinical data to predict suicidality. The results are promising, although we used basic, simpler, and routine biometrics (collected via iPhone and Fitbit), compared with data used in previous research. Studies aiming to predict mental state (short term) have used multiple (self-report) measurements and a wide range of bio sensors [12,15,66].

Strengths, Limitations, and Further Testing

We recognize that our study is limited by the short follow-up period of up to a week; thus, future iterations would need to extend to a longer period of study to explore the time sensitivity of model predictions over varying time windows (eg, predicting current risk vs 1 week out). Short-term risk prediction is difficult because any inference is based on limited data, which means that meaningful signals are lost due to noise from highly variable behaviors [13]. There is promise in improvement as the amount of data available for training and testing increases. Previous research and machine learning literature [67-70] points to expected improvement in performance and reliability in test results as the sample size grows, particularly in this classification setting. We expect that roughly doubling the sample size would achieve more practical results where the possibility of implementation would be appropriate.

Although the results from this feasibility study have been promising in producing a signal, in terms of operationalizing risk for suicide, future steps would be moving beyond survey-generated risk scores. Before taking that leap, the intermediate step would be to further validate the algorithmic results by collecting additional, more substantial test data. Where the experiment excels in the data sources are diverse rather than strictly clinical and allow for natural extension to outpatient settings. In addition, given the probabilistic nature of the algorithm, there will naturally continue to be a trade-off between false positives and false negatives as the model improves, and hence, medical, human attention in decision making will remain critical. We propose that the algorithmic approach provides a supplement and an additional facet to clinical judgment.

Therefore, having achieved a signal from the data for risk in phase 1, phase 2 (proof of concept) will involve collecting more data to not only see if modeling improves but also to test other models such as predicting the risk score trajectory. Enforcing a minimum of 2 C-SSRS assessments, we can try to model changes in risk. We also intend to experiment with more features, particularly those involved in text mining, as most journaling features were relatively surface level. Moreover, we aim to look at prediction stability over time, as this prediction was made within a couple of days from usage to assessment.

Our final aim is to form our own standard so as to break away from dependency on the C-SSRS, as we look to go beyond information gathered in a formal survey that depends solely on human judgment. Further research will enable us to test the viability of automation and machine learning to identify suicide risk by comparing predictions of risk with eventual outcomes as well as testing out the model in different settings and populations (eg, community).

We would also like to point out that, although mobile phones and apps are ubiquitous and have the potential to be an efficient and cost-effective approach to addressing mental health problems [71], this study indicated that there are certain costs that limit the widespread adoption of health apps within mental health services (weather inpatient and community settings). These are related to access to smartphones, connectivity, updating, and maintenance of technology. The premise for this study was that, in line with the UK population statistics indicating that approximately 95% of households own a mobile phone [72], of which a high proportion are smartphones, participants would have access to and use their own smartphones for the study. Following initial scoping, the authors realized that only a small proportion of inpatients had access to a smartphone. In addition, the SWiM app was configured (in its current testing form) to operate only with iOS products, that is, an iPhone. We cannot confirm the extent to which given participants study iPhones might have affected the results; this is something that needs to be further explored. We can, however, highlight that participants were enthusiastic about using Fitbit wearables and the Fitbit app on the phone, which may or may not have encouraged them to use the SWiM app as well.

Conclusions

Although in its early stages, research in this area suggests that using smartphones to enquire about suicidal behaviors can be a valuable approach and not a risk factor for increasing suicidal ideation [12]. Given the heavy reliance on smartphones and mobile devices around the world, this readily available source of data is an important and highly underutilized source that has good potential to improve mental health risk prediction and prevention and advance global mental health. However, although full automation and independence of clinical judgment or input would be a worthy development for those individuals who are less likely to access specialist mental health services, and for providing a timely response in a crisis situation, we need to acknowledge the ethical and legal implications of such advances in the field of psychiatry [72,73]. The use of machine learning in suicide prediction needs a strong evidence base across different settings, populations, suicidal behaviors, and datasets, before considering full integration in health care settings. For the time being, if proven accurate and scalable, machine learning algorithms for suicide risk detection are likely to complement rather than replace clinical judgment [72]. Although smartphones provide us with opportunities to gather data on real-time dynamic risk factors for suicidal behavior, which would be almost impossible to monitor on discharge (from mental health settings), more research is needed to validate the utility of risk markers for suicide behavior and confirm a safe and clinically effective way to use these data to...
inform practice [13]. More work is needed before we can achieve safe and effective integration within mental health settings, while remaining attentive to key ethical implications. An interesting ethical dimension is related to the use of the KNN algorithm, which requires continued access to the pooled data of (at least a subset of) multiple participants to subsequently label new cases. Although testing this in a controlled setting with inpatients who have provided consent for the use of their data might be straightforward, it is uncertain if service users in the community would accept to have their suicidal trajectory data shared for this purpose or how mental health services would be able to bridge the gap. Furthermore, to achieve high accuracy in terms of short-term risk prediction, a wide variety of data from multiple sources will need to be collected, with data integration as a key component [13]. We, therefore, expect multiple data governance, privacy, and intellectual property issues at stake.

Acknowledgments
This work was supported by the Department of Health, Global Digital Exemplar (ref: CENT/DIGEX/RW4/2017-10-16/A). The authors would like to acknowledge Abhijith Nagaraja’s contribution. He was the lead architect and engineer to design and deploy the iOS apps used to collect data for study participants. Abhijith was also responsible for designing and developing the backend software to ensure that data are captured and stored as per privacy policy standards requirements.

Authors’ Contributions
AH has contributed to the design of the study, set up and management of the study, drafted sections of the paper, and revised it critically. GC has conducted all the data analysis and interpretation, drafted sections of the paper, and contributed to the overall revision. AB and AW collected all the data for the work and critically revised the paper. CK contributed to the design of the study, drafted sections of the paper, and critically revised it. RS was responsible for managing the design and development of the software app, product analytics, and delivery as well as revising the paper before submission. DF was the chief investigator for this project, having contributed to the design and oversight of the project as well as drafting sections of the paper and critically reviewing it.

Conflicts of Interest
AH, GC, AB, AW, and RS have nothing to disclose. CK reports grants from the Stanford University School of Medicine, during the conduct of the study, and personal fees from The Risk Authority in 2016 before the study, outside the submitted work. DF reports being a member of the Board of Managers for Innovation Augmented Intelligence Medical Systems Psychiatry, a limited liability company between Mersey Care National Health Service Foundation Trust and The Risk Authority, Stanford. This has overseen the development of technology that is undergoing evaluation in this research study.

Multimedia Appendix 1
Sample training data.
[XLSX File (Microsoft Excel File), 13 KB - mhealth_v8if6e15901_app1.xlsx ]

References


Abbreviations

C-SSRS: Columbia-suicide severity rating scale
SVM: support vector machine
SWiM: Strength Within Me
Abstract

The current pandemic of the coronavirus disease (COVID-19) has highlighted the importance of rapid control of the transmission of infectious diseases. This is particularly important for COVID-19, where many individuals are asymptomatic or have only mild symptoms but can still spread the disease. Current systems for controlling transmission rely on patients to report their symptoms to medical professionals and be able to recall and trace all their contacts from the previous few days. This is unrealistic in the modern world. However, existing smartphone-based GPS and social media technology may provide a suitable alternative. We, therefore, developed a mini-program within the app WeChat. This analyzes data from all users and traces close contacts of all patients. This permits early tracing and quarantine of potential sources of infection. Data from the mini-program can also be merged with other data to predict epidemic trends, calculate individual and population risks, and provide recommendations for individual and population protection action. It may also improve our understanding of how the disease spreads. However, there are a number of unresolved questions about the use of smartphone data for health surveillance, including how to protect individual privacy and provide safeguards against data breaches.

(JMIR Mhealth Uhealth 2020;8(6):e19457)  doi:10.2196/19457

KEYWORDS

COVID-19; China; mobile health; mobile phones; smartphones; contact tracing; social media; spatiotemporal data; GPS; disease tracking; public health; infectious disease; virus

The recent outbreak in China of the coronavirus disease (COVID-19), an infectious disease caused by a new coronavirus (novel coronavirus [2019-nCoV] or severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]), has developed into a global public health crisis. The estimated basic reproductive number \(R_0\) of 2019-nCoV, or the number of additional cases infected by each case, is 2.2 [1]. This is comparable with severe acute respiratory syndrome (\(R_0\) 2.2-3.6) [2] and Middle East respiratory syndrome (\(R_0\) 2.0-2.8) [3], showing the high infectivity of SARS-CoV-2. Early tracing and quarantining of close contacts are critical in cutting off the transmission chain and limiting the scale of any epidemic [4]. The current patient-focused biomedical data-based strategy for epidemic prevention and control focuses on identifying patients by clinical examinations. This assumes that patients will go to a clinical institution for diagnosis as their symptoms worsen and doctors will then trace their contacts, provided that their memories are good enough. However, in a modern society characterized by high-speed mobility and high-dimensional complexity, this patient-focused biomedical data-based strategy does not permit sufficiently prompt tracing and quarantining of potential sources of infection in the population. It can, therefore, fail to halt the
spread of the disease [5]. Up to 84% of patients with COVID-19 are asymptomatic or have only mild symptoms [6], and their diagnosis tends to be delayed. It is therefore difficult, if not impossible, to trace and quarantine all patients with latent disease before they infect others. On March 11, 2020, the World Health Organization officially declared COVID-19 a pandemic [7]. Many countries have limited travel and shut down cities, but unpredictable latent infections could worsen disease transmission and overwhelm local medical and social systems. It is much more cost-effective to cut off the transmission chain and protect the majority of susceptible people from infection than to treat numerous patients in hospitals. Developing an effective strategy for precise individual monitoring that can trace close contacts promptly and protect the majority of people without limiting basic city functions is, therefore, becoming increasingly urgent worldwide.

Smartphones are widely used electronic products that can detect characteristics of users, such as spatiotemporal trajectory and social contacts. They are already used in health care through websites and apps such as WeChat, Twitter, and Facebook, opening up a new field in medical and scientific research known as mobile health. Current practice in mobile health mainly involves medical surveys, chronic disease interventions, and health education around infectious diseases [8]. To harness this technology for infectious disease prevention and control, systems must be developed to exploit the core information needed: personal spatiotemporal trajectory data, which can be gathered using a GPS and geospatial artificial intelligence (GeoAI) technologies. A GPS is a satellite-based radio navigation system that can provide real time geolocation and time information to a GPS receiver anywhere on Earth. GeoAI is the combination of artificial intelligence (AI) and geographic information systems (GIS), and has been used in public health [9]. For example, Google Flu Trends was used to forecast state-level trends of seasonal flu epidemics in the United States [10]. We suggest that improving the resolution of monitoring and forecasting could enable accurate contact tracing and precise individual-level protection, both of which are critical in the management of COVID-19. They are also realizable using existing technologies.

We therefore proposed the spatiotemporal reporting over network and GPS (STRONG) strategy, a system integrating GPS and social media via a smartphone app and GeoAI. STRONG is characterized by dynamic involvement of the whole population, including both diagnosed patients in and out of hospitals, as well as apparently healthy individuals. This is important in control of severe communicable diseases with a high R0 such as COVID-19. Updated spatiotemporal trajectory data from GPS–enabled smartphones can be collected through routinely used social media apps. The back end system analyzes spatiotemporal data from all users and promptly and accurately traces the close contacts of all patients. Unlike current memory-dependent contact tracing methods, spatiotemporal trajectory data in the cloud can provide unique, detailed, permanent, and traceable spatiotemporal trajectory data for both diagnosed patients and healthy individuals. This permits early tracing and quarantine for potential sources of infection. The massive volume of spatiotemporal data collected can also be merged with other data sources including clinical data from medical institutions and macroscopic data such as GIS data from government sources and, if possible, volunteer-interacting data from the smartphone apps ecological system, like location data deriving from Uber history or payment history. These integrated medical, biological, and demographic data can then be analyzed using AI–based methods including GeoAI to predict epidemic trends on a macrolevel [11]. The system will generate an algorithm to calculate time-specific individual and location risks, and guide personal protective measures. It will also provide important reference information to help governments formulate policies for a protective network for the healthy population. Crucial epidemiological data points, such as the infection’s transmissibility through each route and the role played in transmission by subclinical, asymptomatic, and mild infections, could also be calculated, providing a clearer understanding of the spread of the epidemic than has been possible to date [12].

To put STRONG into practice for COVID-19 prevention and control, our team has developed the WeChat mini-program Geo WeChat AI System (Geo-WAS). WeChat is the most popular multifunctional social media app in China, with 1.15 billion monthly active users, or 80% of the Chinese population, and created about 29.63 million job opportunities in 2019 [13]. It is, therefore, the obvious platform for collection of personal spatiotemporal trajectory and social behavior data, especially in the current emergent epidemic situation in China. A key component of the WeChat platform is the technology of mini-programs, or powerful applets within the WeChat ecosystem [14]. Our mini-program collects data from users’ voluntary WeChat activities, including time and location labels, volunteered smartphone assisted real world activities history over the previous 14 days, and the current maximum incubation period for COVID-19, to generate an updated space-time Quick Response (QR) code to use for identification.

We have also created an individual dynamic spatiotemporal risk index to quantify the real time accumulative exposure risk for each user. The risk index is a ratio of the weighted proportion of locations containing infectious, suspected, and pending users defined as the close contacts of individuals who are either infectious or suspected to be in all areas where the user has been within the last 14 days. The spatiotemporal risk index is calculated using the formula:

\[
R(x) = \frac{L_{sp} + L_{cr} + L_{cs}}{L_{sp} + L_{cr} + L_{cs} + L_{xp}}
\]

where \(R(x)\) is the individual dynamic spatiotemporal risk index for an individual \(x\) in Geo-WAS; \(L_{sp}\), \(L_{cr}\), and \(L_{xp}\) are the number of locations that user \(x\) has been to in the past 14 days that contain any marked infectious, suspected, and pending users; and \(L_{cs}\) is the number of locations that user \(x\) has been to over the past 14 days. The risk index uses a spatial transmission trajectory generated for each user from GPS data, rather than calculating an assumed probability of infection during each contact. It, therefore, provides a richer description of epidemic transmission than a traditional susceptible, infected, recovered model. This solves the problem of precise contact tracing.
The STRONG strategy (see Figure 1) will enable accurate location and tracking of sources of infection in the population and offer protection for the most susceptible individuals. In Figure 1, A shows a patient-focused biomedical data-based strategy, which centers on the source of infection and is unable to identify untraceable mild and asymptomatic patients in the population. B and C show the web-like individual tracking network built in STRONG for patients who are infected and the vast majority of the healthy population. D shows that STRONG can collect detailed, unique, permanent, and traceable spatiotemporal trajectory data for each user through social media and GPS-enabled smartphone technology, and use GeoAI algorithms to analyze data in real time. E shows the evolution of communication and transportation technology, resulting in the era of the Internet of Things, with high-speed dynamicity and high-dimensional complexity.

Figure 1. Comparison of patient-centered biomedical data-based strategy and digital data-oriented real time intelligent epidemic prevention and control strategy. GeoAI: geospatial artificial intelligence.

Our mini-program Geo-WAS was published on January 31, 2020. In March 2020, the Chinese government implemented a national monitoring system, using mobile phone base station positioning data to generate health QR codes as proof of an individual’s health condition. This system allowed people with the green QR code to return to work or school and even travel between different areas or provinces; this promoted the resumption of work, helped the precise control and smart management of COVID-19, and successfully prevented further COVID-19 transmission in China. Base station positioning data is not precise enough to distinguish all the risk contact compared with GPS data. However, for GPS data, a major unresolved issue is “noise” and false positive or negative identification; thus, the proper time and distance threshold to identify a risk contact or risk area in a GPS-based system still remains to be tested and adjusted in future practice. Similar applications of mobile health (mHealth) for COVID-19 such as “TraceTogether” from Singapore, “Aarogya Setu” from India, or “Big Sensor Data” use Bluetooth or SMS to help monitor contact in a different way [15-18]. Some technology firms like Apple and Google also cooperated to develop similar Bluetooth-based apps in iOS and Android systems. Compared with these apps still in development or just into use, which regard the smartphone as a simple radar, some mature ecology-formed platforms like WeChat and Alipay in China have already integrated multiple types of data including social communication, mobile payment, online community, and sound privacy agreements, and could, thus, gather traceable data both from the virtual and real world. Though the mHealth-related products have been recently developing quickly, these apps rely on users or health care workers to report suspicious or confirmed COVID-19 cases on the system to update the risk area timely. There are concerns about personal confidentiality and privacy [19]. First, it has become acceptable to authorize the use of personal data including personal information, activity status data, and spatiotemporal data in the use of social media apps and other apps like Uber. However, the tension between proliferation of digital health care data and data privacy concerns is important because of the implications of any data breach or loss of health information. Some recently proposed novel protocols improving data transmission procedure could not satisfy complex data and
further exploration is needed [16,20]. Blockchain technology may be helpful in managing this tension. This is a distributed database that originated from financial instrument research involving cryptocurrencies, of which bitcoin was the first well-established example, and could provide a reliable method for cross-platform management and sharing of medical data while ensuring its confidentiality [21]. Using blockchain technology, cryptographic health and operations data can be transferred from node to node by an automatically generated smart contract, a software protocol for legal agreements. It may also be worth considering people’s acceptance of the data privacy implications of Geo-WAS or the health QR code in light of the threat from COVID-19 to personal and public health. Second, collection, interpretation, and publication of information in a professional and appropriate way remain challenging. Smartphone social media apps are widely used, and it would not be an exaggeration to regard smartphones as essential for a substantial portion of the world’s population. Their use for health surveillance, therefore, goes beyond informing clinical staff and requires cooperation of an interdisciplinary team of biologists, clinicians, data scientists, and engineers. This could help to avoid misunderstandings of results or stigmatizing communities or individuals. Third, the dilemma between individual risk and population benefit is complex and varies in different areas and countries. It is an important part of public health ethics. To protect individuals and countries from the threat of COVID-19 infection and considering the necessity for policy changes that would interrupt personal lives and basic city functions, government should consider the use of STRONG, which will require recruitment of experts, close supervision of the process, and timely feedback.

In summary, STRONG is potentially a powerful strategy for epidemic prevention and control supported by smartphone social media app-based spatiotemporal trajectory data collection, big data compilation from multiple sources, and GeoAI–based real time data analysis. We have developed and published a suitable mini-program, Geo-WAS, and collected data for further analysis. There are, however, a number of ethical and technical challenges remaining, and we have not yet carried out a formal pilot study. However, during the current COVID-19 epidemic, similar health codes from other sources have been used to control the COVID-19 transmission without disrupting the regular social order [22].

Acknowledgments

This program is supported by the Central South University Emergency Project for COVID-19 (Grant No. 13400/160260007) and Innovative R&D Project of Hunan Development and Reform Commission (2019/875) and Progress and Innovation Program of Hunan Provincial Department of Transportation (2018/234 - 201835). We thank Melissa Leffler, MBA, from the Edanz Group for editing a draft of this manuscript and helping to draft the abstract.

Conflicts of Interest

None declared.

References


Abbreviations

AI: artificial intelligence
COVID-19: coronavirus disease
GeoAI: geospatial artificial intelligence
Geo-WAS: Geo WeChat AI System
GIS: geographic information systems
mHealth: mobile health
QR: Quick Response
R0: basic reproductive number
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2
STRONG: spatiotemporal reporting over network and GPS
2019-nCoV: novel coronavirus

©Shaoxiong Wang, Shuizi Ding, Li Xiong. Originally published in JMIR mHealth and uHealth (http://mhealth.jmir.org), 10.06.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,
provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on http://mhealth.jmir.org/, as well as this copyright and license information must be included.
Using eHealth to Support COVID-19 Education, Self-Assessment, and Symptom Monitoring in the Netherlands: Observational Study

Thomas Timmers¹,², MSc; Loes Janssen³, PhD; Joep Stohr⁴, MD; J L Murk¹, MD, PhD; M A H Berrevoets⁴, MD

¹Interactive Studios, Rosmalen, Netherlands
²Radboud University Medical Center, Radboud Institute for Health Sciences, IQ Healthcare, Nijmegen, Netherlands
³Máxima Medical Center, Veldhoven, Netherlands
⁴Elisabeth Twee Steden Hospital, Tilburg, Netherlands

Corresponding Author:
Thomas Timmers, MSc
Interactive Studios
Huisbergenweg 6
Rosmalen, 5249JR
Netherlands
Phone: 31 0650617047
Email: thomas@interactivestudios.nl

Abstract

Background: The coronavirus disease (COVID-19) situation demands a lot from citizens, health care providers, and governmental institutions. Citizens need to cope with guidelines on social interaction, work, home isolation, and symptom recognition. Additionally, health care providers and policy makers have to cope with unprecedented and unpredictable pressure on the health care system they need to manage. By providing citizens with an app, they always have access to the latest information and can assess their own health. This data could be used to support policy makers and health care providers to get valuable insights in the regional distribution of infection load and health care consumption.

Objective: The aim of this observational study is to assess people's use of an app to support them with COVID-19 education, self-assessment, and monitoring of their own health for a 7-day period. In addition, we aim to assess the usability of this data for health care providers and policy makers by applying it to an interactive map and combining it with hospital data. The secondary outcomes of the study were user’s satisfaction with the information provided in the app, perceived usefulness of the app, health care providers they contacted, and the follow-up actions from this contact.

Methods: This observational cohort study was carried out at the nonacademic teaching hospital “Elisabeth Twee Steden” (ETZ) in Tilburg, Netherlands. From April 1, 2020, onwards ETZ offered the COVID-19 education, self-assessment, and symptom tracking diary to their already existing app for patient education and monitoring.

Results: Between April 1 and April 20, 2020, a total of 6194 people downloaded the app. The self-assessment functionality was used abundantly to check one’s health status. In total, 5104 people responded to the question about severe symptoms, from which 242 indicated to suffer from severe symptoms. A total of 4929 people responded to the question about mild symptoms, from which 3248 indicated to suffer from these. The data was successfully applied to an interactive map, displaying user demographics and health status. Furthermore, the data was linked to clinical data. App users were satisfied with the information in the app and appreciated the symptom diary functionality. In total, 102 users reached out to a health care provider, leading to 91 contacts.

Conclusions: Our study demonstrated the successful implementation and use of an app with COVID-19 education, self-assessment, and a 7-day symptom diary. Data collected with the app were successfully applied to an interactive map. In addition, we were able to link the data to COVID-19 screening results from the hospital’s microbiology laboratory. This data could be used to support policy makers and health care providers to get valuable insights in the regional distribution of infection load and health care consumption.

Trial Registration: Netherlands Trial Register NL8501; https://www.trialregister.nl/trial/8501

(JMIR Mhealth Uhealth 2020;8(6):e19822) doi:10.2196/19822
INTRODUCTION

Background

In December 2019, an outbreak of the coronavirus disease (COVID-19) began in Wuhan, China. As of April 25, 2020, the virus had been identified in 185 countries, accounting for 2.4 million confirmed cases and over 17,000 confirmed deaths [1]. On February 26, 2020, the first documented case in the Netherlands was registered at the Elisabeth Twee Steden (ETZ) Hospital in the city of Tilburg [2], and at the time of writing, there were 46,952 confirmed cases and 5,990 confirmed deaths in the Netherlands [3].

The COVID-19 situation demands a lot from citizens, health care providers, and governmental institutions. Citizens, for instance, need to cope with guidelines on social interaction, work, home isolation, and symptom recognition [4]. In a new situation, such as this pandemic, being knowledgeable about what severe symptoms are, how to measure them, and how to act according to the latest guidelines could be considered a challenge for many of them. Additionally, health care providers and policymakers have to cope with unprecedented and unpredictable pressure on the health care system they need to manage.

Electronic health (eHealth) offers a potential powerful means to support all stakeholders in situations like these. By providing citizens with an app, they always have access to the latest trustworthy information and education, and can be actively notified of important changes by a push notification [5]. Furthermore, the app allows them to track their symptoms and gather valuable data for both the users themselves, as well as health care providers and policy makers, who can use this data on a more aggregated level to assess the local or regional health status and (expected) pressure on the health care system [6-10]. The effective use of eHealth has been demonstrated before in the management of chronic diseases and other treatments [11-15].

Objectives

The aim of this observational cohort study is to assess people’s use of an app to support them with COVID-19 education, self-assessment, and monitoring of their own health for a 7-day period by using a symptom diary. In addition, we aimed to assess how this data would be useful for health care providers and policymakers by applying it to an interactive map and combining it with hospital data.

METHODS

Study Design

This observational cohort study was carried out at the nonacademic teaching hospital ETZ in Tilburg, Netherlands. On April 1, 2020, ETZ added the COVID-19 education, self-assessment, and symptom tracking diary to their already existing app for patient education and monitoring, called ETZ.
(RIVM; Rijksinstituut voor Volksgezondheid, Welzijn en Sport) and a link to RIVM’s website for the latest information. Changes in the RIVM’s mitigation strategies were updated in the app, to assure users always had the same guidelines available. Before users continued, they were informed about the fact that the information in the app might not apply to them or that the information in the app cannot replace the personal advice they receive from a health care professional.

Next, users were requested to answer three questions that would provide the hospital with information about the users themselves (age and gender) and their geographical location (4 digits of their postal code) to assess whether geolocation data could be used during outbreaks such as with COVID-19.

After these introductory steps, the app provided a choice between six topics: I want to check my health, I have an appointment in the hospital, I want to visit or accompany a patient, I am in isolation at the hospital, I am in isolation at home, and I want general information about COVID-19 (Figure 1).

The “I want to check my health” part of the app offered users the possibility to self-assess their current COVID-19 health status, focusing on severe symptoms (body temperature of 38°C or more, in combination with shortness of breath), mild symptoms (sore throat, coughing, rise in temperature from 37.5°C to 38°C, runny nose, sneezing, and diarrhea), and underlying diseases such as chronic obstructive pulmonary disease, asthma, diabetes, or heart failure. Based on the input the user provided, an outcome and advice on how to handle it was presented. This decision tree was based on the guidelines from the health authorities.

After the completion of the self-assessment, users were provided the possibility to keep track of their symptoms (body temperature and shortness of breath) for a 7-day period. When they chose to do so, they were requested to fill in the date of the first day of symptom tracking and were educated on when and how to measure and report the symptoms. They were informed once again that there was no connection between the app and the hospital to make sure users did not expect a response from ETZ health care staff. Users received daily push notifications at 10 AM to report their body temperature and at 2 PM to report their shortness of breath.

The other five parts of the app focused on educational content. Patients who wanted information about their appointment in the hospital got hospital specific information about planned consultations and surgeries. Patients who wanted to visit or accompany a patient in the hospital were provided the local guidelines on, for instance, visiting hours and the maximum number of visitors or people that could accompany a patient. Information about isolation in the hospital came from the hospital itself whereas the information on self-isolation and general COVID-19 information came from the RIVM’s website. All information in the app was presented in Dutch.

**Figure 1.** Examples of the coronavirus disease guide in the Elisabeth Twee Steden Behandelwijzer app (in Dutch). From left to right: the welcoming of patients to the app (including a video of the intensive care unit), the main menu to choose the type of information or functionality, part of the self-assessment tool (mild symptoms and underlying diseases are displayed), and the result of the 7-day tracking of symptoms (progression of body temperature is displayed).

**Study Outcomes**

The primary outcome of the study was the number of users that completed the COVID-19 self-assessment and the number of users that used the symptom diary. In addition, we aimed to assess the usability of this data for health care providers and policy makers by applying it to an interactive map and combining it with hospital data. The data users provided was divided into three regions, based on postal code regions: Tilburg City Area, Greater Tilburg Region, and other. The secondary outcomes of the study were user’s satisfaction with the information provided in the app, perceived usefulness of the app, health care providers they contacted, and the follow-up
actions from this contact. All outcomes were measured by self-reported in-app questionnaires (Textbox 1).

Study outcomes were measured at baseline (user characteristics), after users indicated to want to self-assess their health on a daily basis for 7 days after using the health monitor (body temperature and shortness of breath), and on the seventh day after the start of the symptom diary to assess satisfaction, added value, contact with health care providers, and follow-up actions (Table 1).

Textbox 1. Overview of used questions per outcome.

<table>
<thead>
<tr>
<th>User characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• What is your gender (Female, Male, other)?</td>
<td></td>
</tr>
<tr>
<td>• What is your age (numerical value between 18 and 110)?</td>
<td></td>
</tr>
<tr>
<td>• What are the first 4 digits of your postal code (numerical value)?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Do you suffer from a combination of fever (body temperature of 38.0°C or more) and shortness of breath (yes/no)?</td>
<td></td>
</tr>
<tr>
<td>• Do you suffer from any of these mild symptoms: sore throat, mild cough, raise in body temperature (between 37.5 and 38.0°C Celsius), runny nose, sneezing, or diarrhea (yes/no)?</td>
<td></td>
</tr>
<tr>
<td>• Do you have any underlying diseases such as chronic obstructive pulmonary disease, asthma, diabetes, or heart failure (yes/no)?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Body temperature (based on the national RIVM guidelines for COVID-19 symptom screening)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Numeric value, one decimal, ranging from 35.0 to 45.0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shortness of breath (based on the national RIVM guidelines for COVID-19 symptom screening)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Numeric Rating Scale (NRS), 0-10 scale; 0 stands for “no trouble breathing at all,” 5 stands for “some troubles,” and 7 or higher stands for “trouble breathing even when in bed or on a couch.”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Satisfaction with information in the app (self-developed questions, specifically for the study)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• How satisfied are you with the information in this app? NRS, 0-10 scale; 0 stands for “not satisfied at all” and 10 stands for “extremely satisfied.”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Added value of symptom monitoring using the app (self-developed questions, specifically for the study)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Do you find it valuable to monitor your health through this app? NRS, 0-10 scale; 0 stands for “not valuable at all” and 10 stands for “super valuable.”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contact with health care providers (self-developed questions, specifically for the study)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Did you contact a health care provider in the last 7 days? Multiple choice question: no contact with health care providers, contact with general practitioner, contact with hospital, or contact with general practitioner and hospital.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Follow-up actions after contact with health care provider (self-developed questions, specifically for the study)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Was there a follow-up action as a result of your contact with the health care provider? Multiple choice question: none, visit to general practitioner, visit to hospital/emergency department, admittance to hospital, or admittance to intensive care unit.</td>
<td></td>
</tr>
</tbody>
</table>
Study Size
We did not perform an a priori sample size calculation but rather looked at the number of downloads the app had in other hospital regions where it was used. In those regions, the number of downloads was between 1000 and 2000 downloads per region. Given the population size in the Tilburg area and the press release ETZ was planning, we aimed at approximately 2000 downloads.

Statistical Methods
Categorical variables are presented as numbers and percentages. Continuous variables are presented as means (with SD). There was no statistical comparison between groups, since the aim of this study is to assess the use of the app and the usability of the data that comes from it, rather than demonstrating in-between group differences. Descriptive statistics will be presented at postal code level. Data from the health self-assessment will be presented in tabular format. Data from the symptom tracking diary (body temperature and shortness of breath) will be represented in tabular format as well as line graphs. Data plotting for geolocation purposes was performed using raw data and the Google Maps and Google Geolocation application programming interface (API; Google). All data was analyzed using IBM SPSS Statistics for Macintosh, version 25.0 (IBM Corp).

Results
Study Sample
Between April 1 and April 20, 2020, a total of 6194 people used the app (5698 new downloads [92%], 496 existing app users [8%]). The average number of push notifications sent per day during this period increased from 113 to 561 (496%) and the average number of views per day increased from 178 to 611 (343%). The functionality that was used most often was the health self-assessment, which was used by 5326 users (86%). Figure 2 provides an overview of the functionalities of the app and the number of users that used them.

User Characteristics
Of the 6194 downloads, 5364 (87%) users of the app shared data about their gender, and 5328 (86%) shared data about their age. The average age was 50.87 years (SD 14.38), and 46% (n=2455) of users were male. Within the group of users that checked their health (n=4655), the average age was 50.60 years, and 47% (n=2198) of users were male (Table 2). Related to postal codes, a total of 5872 users (85%) reported the first 4 digits of their postal code. Postal code segments were used to divide the app users into 3 groups: Tilburg City (n=1464), Tilburg Region (n=1459) and Other (n=2949).

Table 1. Overview of outcomes per measurement.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Baseline</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>User characteristics</td>
<td>✓</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Self-assessment/quick scan</td>
<td>✓</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Body temperature</td>
<td>N/A</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>N/A</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Satisfaction with information</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
</tr>
<tr>
<td>Added value of symptom monitoring</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
</tr>
<tr>
<td>Contact with health care providers</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
</tr>
<tr>
<td>Follow-up actions after contact with health care provider</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>✓</td>
</tr>
</tbody>
</table>

aNot applicable.
Table 2. User characteristics.

<table>
<thead>
<tr>
<th>Group and characteristic</th>
<th>User</th>
<th>Overall (n=6194)</th>
<th>Users that checked their health (n=5326)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n</td>
<td>5364</td>
<td>2455 (46)</td>
<td>2198 (47)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td></td>
<td>2884 (54)</td>
<td>2436 (53)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td></td>
<td>50.87 (14.38)</td>
<td>50.60 (14.42)</td>
</tr>
<tr>
<td>Age (n=5328), mean (SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Primary Outcomes

Use of the Self-Assessment Functionality

The self-assessment functionality was used abundantly to check user’s health status. In total, 5154 people responded to the question about severe symptoms, from which 242 (4.7%) indicated that they suffered from severe symptoms. A total of 4929 people responded to the question about mild symptoms, from which 3248 (65.9%) indicated that they suffered from these. Finally, 2929 people responded to the question concerning underlying diseases, and 1099 (22.3%) of them indicated to suffer from them. There were only small differences between the results reported by inhabitants of Tilburg City, the Tilburg Region, and the other users (Table 3).

Table 3. Self-assessment results.

<table>
<thead>
<tr>
<th>Region</th>
<th>Severe symptoms</th>
<th>Mild symptoms (yes/no)</th>
<th>Underlying diseases (yes/no)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tilburg City, n/N (%)</td>
<td>53/1178 (4.5)</td>
<td>751/1146 (65.5)</td>
<td>293/1144 (25.6)</td>
</tr>
<tr>
<td>Tilburg Region, n/N (%)</td>
<td>40/1111 (3.6)</td>
<td>664/1055 (62.9)</td>
<td>229/1055 (21.7)</td>
</tr>
<tr>
<td>Other, n/N (%)</td>
<td>149/2865 (5.2)</td>
<td>1833/2727 (67.2)</td>
<td>577/2721 (21.2)</td>
</tr>
<tr>
<td>Total, n/N (%)</td>
<td>242/5154 (4.7)</td>
<td>3248/4928 (65.9)</td>
<td>1099/4920 (22.3)</td>
</tr>
</tbody>
</table>

Use of the 7-Day Symptom Monitoring Diary

The 7-day symptom diary was initially used by 1378 people. The data provided by users in the three groups is consistent in terms of the outcome (body temperature or shortness of breath) and standard deviation. There was a decrease in the number of users that completed the body temperature diary for 7 consecutive days, but this decrease was more pronounced in users from the “Other” area (642 on day one compared to 137 [21%] on day seven) than it was in Tilburg City (378 on day one compared to 119 [31%] on day seven) or Tilburg Region (338 on day one compared to 116 [34%] on day seven; Figure 3; see Multimedia Appendix 1 for all body temperature data). The same pattern is demonstrated in the shortness of breath diary, where the decrease was also more pronounced in users from the “Other” area (405 on day one compared to 83 [21%] on day seven) than it was in Tilburg City (239 on day one compared to 85 [35%] on day seven) or Tilburg Region (212 on day one compared to 72 [34%] on day seven).

In the “Other” group, there were 405 users that completed the seventh day of the diary on day one compared to 83 (20%) on day seven of all cases, compared to 239 users on day one to 85 (32%) on day seven in Tilburg City, as well as in the Tilburg Region (Figure 4; see Multimedia Appendix 1 for all shortness of breath data).
Figure 3. Body temperature results per day over a 7-day period as reported by inhabitants of Tilburg City, the Tilburg Region, and other areas.
Usability of the Data for Health Care Providers and Policy Makers

The data that was gathered during the use of the app focuses on 4 domains: demographic information (age and gender), geographical information (postal code), COVID-19 symptoms, and contact with health care providers. Combining this data creates an overview of symptoms per postal code, to which clinical data from the ETZ about patients that have been tested positive for COVID-19 can be added (Table 4).

Moreover, the data can be used to create a (near) real time COVID-19 map for the city of Tilburg by applying the data to a Google Maps overview through the Google Maps API (Google; Figure 5).

Table 4. An example of combined user data, health status, and contact with health care providers for the postal codes with the highest number of app users.

<table>
<thead>
<tr>
<th>Postal code</th>
<th>Users, n</th>
<th>Age, mean (SD)</th>
<th>Underlying diseases, n</th>
<th>Mild symptoms, n</th>
<th>Severe symptoms, n</th>
<th>Tested positive(^a), n</th>
<th>General practitioner, n</th>
<th>Contact Follow-up</th>
<th>Emergency department, n</th>
<th>Contact Follow-up</th>
<th>Hospital, n</th>
<th>Contact Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>5045</td>
<td>174</td>
<td>52.74 (11.51)</td>
<td>37</td>
<td>89</td>
<td>4</td>
<td>12</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5038</td>
<td>101</td>
<td>54.08 (15.52)</td>
<td>14</td>
<td>49</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5021</td>
<td>90</td>
<td>49.00 (15.55)</td>
<td>22</td>
<td>49</td>
<td>4</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5046</td>
<td>88</td>
<td>48.42 (13.20)</td>
<td>18</td>
<td>49</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5011</td>
<td>81</td>
<td>52.59 (17.86)</td>
<td>17</td>
<td>35</td>
<td>5</td>
<td>21</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\)Clinical data from the Elisabeth Twee Steden microbiology laboratory (date range: April 1 to April 20, 2020).
Secondary Outcomes

Satisfaction With the Information Provided Through the App

In total, 718 users answered the question about their satisfaction with the information in the app. Users in all three groups indicated that they were very satisfied with the information (mean 7.93, SD 1.60 in Tilburg City; mean 7.86, SD 1.51 in the Tilburg Region; and mean 8.08, SD 1.49 in the other area).

Added Value of Symptom Monitoring Through the App

In total, 671 users answered the question about the added value of monitoring their symptoms by means of an app. Users in all three groups indicated the added value to be high (mean 8.04, SD 1.83 in Tilburg City; mean 7.95, SD 1.65 in the Tilburg Region; and mean 8.14, SD 1.62 in the other area).

Contact With Health Care Providers and Follow-Up Actions Performed

In total, 638 users answered the question about any COVID-19–related contact they had with a health care provider in the 7 days after the start of the symptom diary in the app. Overall, 84% of users (n=526) reported not to have contacted a health care provider. In cases where users did report reaching out, 86 out of 102 contacts (84%) were initiated by users reporting severe (n=2) or mild symptoms (n=84). General practitioners were contacted most frequently (n=92), followed by the hospital (n=8) and the emergency department (n=2). On average, 87.75% of all contact with a health care provider led to a physical appointment with that provider (Table 5).

Table 5. Contact with health care provider.

<table>
<thead>
<tr>
<th>Region</th>
<th>Users, n</th>
<th>General practitioner</th>
<th>Emergency department</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contacts, n</td>
<td>Visits, n (%)</td>
<td>Contacts, n</td>
<td>Visits, n (%)</td>
</tr>
<tr>
<td>Tilburg City</td>
<td>172</td>
<td>21</td>
<td>20 (95)</td>
<td>1</td>
</tr>
<tr>
<td>Tilburg Region</td>
<td>174</td>
<td>25</td>
<td>22 (88)</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>284</td>
<td>46</td>
<td>38 (83)</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>638</td>
<td>92</td>
<td>80 (87)</td>
<td>2</td>
</tr>
</tbody>
</table>
Discussion

Principal Findings

The results of our study demonstrate the effect and use of an app to provide users with COVID-19 education and functionalities for self-management and symptom management through a 7-day diary. With regards to the primary outcome, over 6000 users downloaded the app in the first 2 weeks after it became available through the app stores. Over 5000 users used the self-assessment tool and over 1300 users started the symptom diary. In addition, over 5200 users were willing to share their characteristics (age and gender) and geographical data (first 4 digits of their postal code). Furthermore, the app received positive evaluations related to the satisfaction with the information and the added value of keeping track of symptoms. Next to this, the app provided valuable insights into the care users had consumed by means of contacting health care providers, as well as the follow-up visits to these providers. Finally, the data that was gathered through the app, could be applied to an interactive map, displaying the health status and characteristics of users, and could be associated with the clinical data from the hospital on a postal code level. Even though the numbers in this study are too small, future data might indicate an association between the number of self-reported severe symptoms and the number of people that are actually tested positive for COVID-19. Therefore, self-reported data could help to gain insights with regard to the spread of the virus in a certain region.

To our knowledge, our study is the first to assess the effectiveness of this combination of education (both national and local hospital-specific guidelines) and self-assessment and symptom tracking functionality. We consider this combination of features, content, and the trusted regional health care provider not only as a major strength of the study but also as a truly differentiating factor of the app itself, especially in these challenging times, where according to the World Health Organization (WHO) “we are not just fighting an epidemic, but an infodemic as well.” The WHO defines infodemics as an excessive amount of information about a problem, which makes it difficult to identify a solution (and allows for misinformation, disinformation, and rumors to make its entrance during a health emergency). The number of downloads from outside the Tilburg Region clearly demonstrates people’s need for apps with these kinds of functionalities. Many of the apps that are used during this pandemic focus on just one aspect, for instance data collection through symptom tracking—sometimes even without providing feedback to users on what the data means or allowing them to track their progress. Another strength of the study was demonstrated by the number of people that were willing to enter their (nonmandatory) personal data into the app, even though it was clearly stated that the app was not connected to the hospital.

A limitation on the use of the app is the decline in the number of users that continued using the symptom diary during the course of the week. Such reductions in compliance have been reported before and often have to do with user’s missing out on personalized features such as reminders, coaching, and interpretation of the data (and the relation to short- and long-term goals) [17]. Within the app we did implement reminders, notifications, and direct feedback on the data users provided about body temperature and shortness of breath. This feedback, however, consisted of generic messages such as “your body temperature is within a safe range” and is something that needs to be tailored to a patient’s specific context. Moreover, measuring body temperature demands some effort and is not as easy to answer than questions like “do you have a runny nose?” or “do you cough?” Using a smart device with sensor data could overcome this barrier [18,19]. The fact that about 700 users reported to be satisfied with the information in the app (mean 8 out of 10 score) and rated the symptom diary with a mean 8 out of 10 score as well demonstrated user’s willingness to participate in this kind of functionality.

Another limitation could be the fact that the app was really a self-assessment tool; the data users provided was not shared with health care professionals. Not linking the app to health care providers was a request from the health care providers themselves, as the app could lower the pressure on the health care system by guiding users in their self-assessments and symptom tracking, and urging them to only contact health care facilities when indicated. A final limitation is the fact that we are not sure about the effect that the education and feedback in the app had on user’s reaching out, or not, to health care providers. In other words, who was reassured by the app and did not contact health care providers, and who was triggered by the app to contact one? In both cases the question remains, was this the right thing to do?

Clinical Implications

Our study demonstrates that eHealth can be implemented rapidly to successfully support people with important and trustworthy information, self-assessment, and symptom tracking functionality. In this case, the project was initiated by the COVID-19 pandemic, bringing together health care providers and a technology provider to create the app. The results from our study might be applied in other countries that are affected by the COVID-19 pandemic as well. Another application could be the use of the app to support people who are recovering from their hospitalization or home isolation, as many of them have suffered in terms of physical or mental fitness. The app could provide trustworthy education, instructions, and exercises to support users. In addition, in this scenario, health care providers and governments benefit as well, as the data that is being captured provides information about the individual patient’s recovery as well as group-level data on their progress. Finally, in future outbreaks of viruses, apps like these could be a valuable solution to share information and gather data to support the people, health care providers, and policy makers.

A next step in the development of solutions like these, could be more personalized information toward the users. In the COVID-19 case, for instance, users can get specific information that is related to their own health based on their own symptoms or underlying diseases, in combination with the health status (symptoms) in their postal code. From a health care provider’s and policy maker’s perspective this data can be valuable as well, as it could predict the pressure on the (local) health care facilities.
or the possible sources of infection, based on real time data instead of near real time or even older information from health care provider electronic health records. Even though our study demonstrated that only a small set of anonymous data is enough to start to add value for all stakeholders, linking to more specific data (for instance personal health records or laboratory data) would enable providing an even more personalized advice and monitoring of outcomes over time while taking into account all necessary privacy and security measures. Of course, the real value of the app and the data depends on the number of people that download and use it; the bigger the portion of the population, the better the data represents them.

**Future Research**

An important aspect of future research could be the user’s needs, especially when it comes down to their willingness to use the symptom diary. Diaries like these have been successfully applied in apps for patients with chronic diseases, in which patients reported the feeling of safety that their own physician was linked to the data [20]. In addition, the data to be collected should be discussed between health care providers and users; what is the data that health care providers need (and in what frequency) for a reliable advice versus what is the data that users are willing to share from a privacy or workload perspective? The question is, what is the optimal balance? Finally, the requirements for the graphical presentation of the data should be assessed by taking into account the preferences from, for instance, health care providers, policy makers, and public health institutions; whereas, the data itself might be valuable to each of the stakeholders, but the preferences for visualization might differ.

**Conclusion**

Our study demonstrated the successful implementation and use of an app with COVID-19 education, self-assessment, and a 7-day symptom diary. Overall, app users were satisfied with the information supplied through the app and appreciated its functionality. Data collected with the app were successfully applied to an interactive map, displaying postal code–specific demographics, health status, and health care consumption. In addition, we were able to link the data to COVID-19 screening results from the hospital’s microbiology laboratory, indicating an association between app users reporting severe symptoms and the number of patients that were tested positive for COVID-19 in the lab. This data could be used to support policy makers and health care providers to get valuable insights in the regional distribution of infection load and health care consumption.

**Acknowledgments**

We would like to thank all the hospital staff that supported us in the development of the content of the app, both in the initial stage as well as later on when new insights were gained. Contributions of the following centers or parties have led to the success of this study: ETZ, Haga Hospital, Noordwest Ziekenhuis, Treant, Ikazia, Orthopedie Eindhoven, VieCuri, Elkerliek, Zuyderland, Laurentius Hospital, Endometriose in Balans, Martini Hospital, Trauma center West, and Airjine. Furthermore, we would like to thank the RIVM for creating valuable content and enabling us to disseminate this through our app. We would also like to thank the team at Interactive Studios as well, as they have worked day and night to make sure the app was ready for use within 1 week. Finally, we would like to thank everyone who used the app and was kind enough to share their data with us.

**Authors’ Contributions**

TT and LJ conceived the study and designed the trial. TT managed the data. TT and LJ provided statistical analysis. TT, LJ, JS, JLM, and MB drafted the manuscript and contributed to its revision.

**Conflicts of Interest**

The principal investigator, TT, is one of the cofounders of Interactive Studios. Interactive Studios is the company that developed the app used in this study. Interactive Studios offered the app used in this study free of charge. The coauthors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest. Moreover, all authors have completed the International Committee of Medical Journal Editors’ uniform disclosure form and declare the following: no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

**Multimedia Appendix 1**

Body temperature 7-day monitoring results and 7-day shortness of breath monitoring results.

**References**


8. NTVJ. De eerste weken van de "OLVG corona check" (The first week of the OLVG Corona Check). Nederlands Tijdschrift voor Geneeskunde 2020;164(C4501):1.
16.RCT: randomized controlled trial

Abbreviations
API: application programming interface
COVID-19: coronavirus disease
eHealth: electronic health
ETZ: Elisabeth Twee Steden
RIVM: Dutch Health Authority
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology