JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 6 (2018), Issue 1 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Reviews

Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review (e7)
Huan Chen, Yanling Chai, Le Dong, Wenyi Niu, Puhong Zhang. 4

The Impact of mHealth Interventions: Systematic Review of Systematic Reviews (e23)
Milena Marcolino, João Oliveira, Marcelo D'Agostino, Antonio Ribeiro, Maria Alkmim, David Novillo-Ortiz. 16

Original Papers

Smartphone App–Based Assessment of Gait During Normal and Dual-Task Walking: Demonstration of Validity and Reliability (e36)
Brad Manor, Wanting Yu, Hao Zhu, Rachel Harrison, On-Yee Lo, Lewis Lipsitz, Thomas Travison, Alvaro Pascual-Leone, Junhong Zhou. . . . 2 7

A Breastfeed-Promoting Mobile App Intervention: Usability and Usefulness Study (e27)
Chih-Jau Wang, Pimwadee Chaovalit, Suporn Pongnumkul. 39

Tanzania Health Information Technology (T-HIT) System: Pilot Test of a Tablet-Based System to Improve Prevention of Mother-to-Child Transmission of HIV (e16)
Sheana Bull, Deborah Thomas, Elias Nyanza, Sospatro Ngallaba. 55

Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry (e10)
Marcin Masalski, Tomasz Grysi ski, Tomasz Kr cicki. 66

Translation and Validation of the Nomophobia Questionnaire in the Italian Language: Exploratory Factor Analysis (e24)
Mohammad Adawi, Nicola Bragazzi, Lidia Argumosa-Villar, Joan Boada-Grau, Andreu Vigil-Colet, Caglar Yildirim, Giovanni Del Puente, Abdulla Watad. 78

Development and Evaluation of a Mobile Personalized Blood Glucose Prediction System for Patients With Gestational Diabetes Mellitus (e6)
Evgenii Pustozerov, Polina Popova, Aleksandra Tkachuk, Yana Bolotko, Zafar Yuldashev, Elena Grineva. 87

Young People's, Parents', and Professionals' Views on Required Components of Mobile Apps to Support Self-Management of Juvenile Arthritis: Qualitative Study (e25)
Jennifer Waite-Jones, Rabiya Majeed-Ariss, Joanna Smith, Simon Stones, Vanessa Van Rooyen, Veronica Swallow. 103

Importance of Active Participation in Obesity Management Through Mobile Health Care Programs: Substudy of a Randomized Controlled Trial (e2)
Bumjo Oh, Ga-Hye Yi, Min Han, Jong Kim, Chang Lee, Belong Cho, Hee Kang. 114

Face-to-Face Versus Mobile Versus Blended Weight Loss Program: Randomized Clinical Trial (e14)
Emalie Hurkmans, Christophe Matthys, An Bogaerts, Leonie Scheys, Karlien Devloo, Jan Seghers. 127

A Wearable Sensor-Based Exercise Biofeedback System: Mixed Methods Evaluation of Formulift (e33)
Martin O'Reilly, Patrick Slevin, Tomas Ward, Brian Caulfield. 138

Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study (e17)
Florence Jimoh, Elizabeth Lund, Linda Harvey, Catherine Frost, W Lay, Mark Roe, Rachel Berry, Paul Finglas. 152

Evaluating Machine Learning–Based Automated Personalized Daily Step Goals Delivered Through a Mobile Phone App: Randomized Controlled Trial (e28)
Mo Zhou, Yoshimi Fukuoka, Yonatan Mintz, Ken Goldberg, Philip Kaminsky, Elena Flowers, Anil Aswani. 166

Using Mobile Health Intervention to Improve Secondary Prevention of Coronary Heart Diseases in China: Mixed-Methods Feasibility Study (e9)
Shu Chen, Enying Gong, Dhruv Kazi, Ann Gates, Rong Bai, Hua Fu, Weixia Peng, Ginny De La Cruz, Lei Chen, Xianxia Liu, Qingjie Su, Nicolas Girerd, Kamilu Karaye, Khalid Alhabib, Lijing Yan, JD Schwalm. 183

A Smartphone App (BlueIce) for Young People Who Self-Harm: Open Phase 1 Pre-Post Trial (e32)
Paul Stallard, Joanna Porter, Rebecca Grist. 198

A Mobile App to Screen for Neurocognitive Impairment: Preliminary Validation of NeuroScreen Among HIV-Infected South African Adults (e5)
Reuben Robbins, Hetta Gouse, Henry Brown, Andries Ehlers, Travis Scott, Cheng-Shiun Leu, Robert Remien, Claude Mellins, John Joska. . 2 0 7

Toward mHealth Brief Contact Interventions in Suicide Prevention: Case Series From the Suicide Intervention Assisted by Messages (SIAM) Randomized Controlled Trial (e8)
Sofian Berrouiguet, Mark Larsen, Catherine Mesmeur, Michel Gravey, Romain Billot, Michel Walter, HUGOPSY Network, Christophe Lemey, Philippe Lenca. 220

Usage of an Exercise App in the Care for People With Osteoarthritis: User-Driven Exploratory Study (e11)
Dorthe Danbjørg, Allan Villadsen, Ester Gill, Mette Rothmann, Jane Clemensen. 229

Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability (e3)
Gitanjali Das, Cynthia Cheung, Camille Nebeker, Matthew Bietz, Cinnamon Bloss. 242

Hospital-Owned Apps in Taiwan: Nationwide Survey (e22)
Hao-Yen Liu, Wui-Chiang Lee, Ying-Chou Sun, Jun-Jeng Fen, Tzeng-Ji Chen, Li-Fang Chou, Shinn-Jang Hwang. 254

Improving Refill Adherence in Medicare Patients With Tailored and Interactive Mobile Text Messaging: Pilot Study (e30)
Rena Brar Prayaga, Erwin Jeong, Erin Feger, Harmony Noble, Magdalen Kmiec, Ram Prayaga. 264

Prevalence of Health App Use Among Older Adults in Germany: National Survey (e26)
Peter Rasche, Matthias Wille, Christina Bröhl, Sabine Theis, Katharina Schäfer, Matthias Knobe, Alexander Mertens. 274

Health Information Technology Usability Evaluation Scale (Health-ITUES) for Usability Assessment of Mobile Health Technology: Validation Study (e4)
Rebecca Schnall, Hwayoung Cho, Jianfang Liu. 285

Concussion Assessment With Smartglasses: Validation Study of Balance Measurement Toward a Lightweight, Multimodal, Field-Ready Platform (e15)
Joseph Salisbury, Neha Keshav, Anthony Sossong, Ned Sahin. 296

Physical Activity Assessment Using an Activity Tracker in Patients with Rheumatoid Arthritis and Axial Spondyloarthritis: Prospective Observational Study (e1)
Charlotte Jacquemin, Hervé Servy, Anna Molto, Jérémie Sellam, Violaine Foltz, Frédérique Gandjbakhch, Christophe Hudry, Stéphane Mitrovic, Bruno Fautrel, Laure Gossec. 308

Corrigenda and Addenda

Correction: Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy (e20)
Yuan Wu, Xun Yao, Giacomo Vespasiani, Antonio Nicolucci, Yajie Dong, Joey Kwong, Ling Li, Xin Sun, Haoming Tian, Sheyu Li. 323

Review

Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review

Huan Chen[1*], MD, MSc; Yanling Chai[1,2*], BS; Le Dong[1,3], BS, Med; Wenyi Niu[2], BS, Med; Puhong Zhang[1,4], PhD

[1]The George Institute for Global Health, Peking University Health Science Center, Beijing, China

[2]School of Public Health, Peking University Health Science Center, Beijing, China

[3]School of Public Health, Hebei Medical University, Shijiazhuang, China

[4]Faculty of Medicine, University of New South Wales, Sydney, Australia

[*]these authors contributed equally

Corresponding Author:
Puhong Zhang, PhD
The George Institute for Global Health
Peking University Health Science Center
Level 18, Horizon Tower B
No. 6 Zhichun Rd
Beijing,
China
Phone: 86 1082800577 ext 512
Fax: 86 1082800177
Email: zpuhong@georgeinstitute.org.cn

Abstract

Background: The application of mobile health (mHealth) technology in reproductive, maternal, newborn, and child health (RMNCH) is increasing worldwide. However, best practice and the most effective mHealth interventions have not been reviewed systematically.

Objective: A systematic review and meta-analysis of studies of mHealth interventions for RMNCH around the world were conducted to investigate their characteristics as well as the features and effectiveness of mHealth interventions.

Methods: Studies of mHealth interventions for RMNCH between January 2011 and December 2016 were retrieved from 6 databases (PubMed, EMBASE, Global Health, China National Knowledge Infrastructure, VIP Database for Chinese Technical Periodicals, and Wanfang Data Knowledge Service Medium). Comparable studies were included in a random-effects meta-analysis for both exclusive breastfeeding (EBF) and antenatal checks (ANC). Descriptive analyses were conducted for mHealth studies with a range of study designs.

Results: Analyses of 245 studies were included, including 51 randomized controlled trials (RCTs). Results showed that there are increasing numbers of studies on mHealth interventions for RMNCH. Although 2 meta-analysis, one with 2 RCTs on EBF (odds ratio [OR] 2.03, 95% CI 1.34-3.08, I^2=25%) and the other with 3 RCTs on ANC (OR 1.43, 95% CI 1.13-1.79, I^2=78%), showed that mHealth interventions are more effective than usual care, almost half (43%) of RCTs showed negative or unclear results on mHealth interventions. Functions described in mHealth interventions were diverse, and the health stages covered were broad. However, single function or single stage appeared to be dominant among mHealth interventions compared with multiple functions or stages.

Conclusions: More rigorous evaluations are needed to draw consistent conclusions and to analyze mHealth products with multiple functions, especially those popular in the app markets.

(JMIR Mhealth Uhealth 2018;6(1):e7) doi:10.2196/mhealth.8998

KEYWORDS

telemedicine; maternal health; child health

Introduction

Reproductive, maternal, newborn, and child health (RMNCH) has improved dramatically in the past two decades, according to a World Health Organization (WHO) report in 2015 [1]. However, there are new challenges, partly because of the changing burden of diseases (such as the increasing prevalence of noncommunicable diseases [NCDs]). Meanwhile, the needs of RMNCH, such as control of infectious disease and ensuring a safe pregnancy, continue to be relevant. Issues such as limited resources and engagement of patients in their health management remain challenges for the improvement of RMNCH health services worldwide [2-5].

In recent years, the rapid development of information and communication technologies (ICTs) in health care worldwide has led to the development of *mobile health (mHealth)* and enabled substantial change in the provision of health services [6,7]. mHealth technology is well suited to designing a patient-centered health service that increases the role of patients in medical treatment and encourages a degree of self-management, which is particularly important for a long-term chronic condition. Moreover, the widespread availability of ICT infrastructure in resource-limited settings provides access to high-quality health information and, in general, requires less staff and specialized health professionals [6,8]. A large number of RMNCH mobile applications (App), sensors, and wearable devices have been developed recently and are currently on the market, with diverse functions, ranging from bio-data monitoring to decision-making assistance [9-12].

Along with the boom of mHealth interventions in RMNCH, the number of studies to describe the development and evaluation of individual interventions is increasing. However, they are not yet extensive enough to provide adequate information to health professionals in making informed decisions about the best apps for particular health issues and situations. The existing systematic reviews either tend to focus on the effectiveness of mHealth interventions in the developing world, such as reviews by Lee et al [13], Sondaal et al [14], and Dahdah et al [15], or they focus on a single condition, such as a study on mHealth interventions for psychiatric conditions in children by Archangeli et al [16]. It seems that no study yet has described the features of RMNCH-related mHealth interventions comprehensively. Therefore, to meet this need, we conducted a systematic review and a meta-analysis of studies of RMNCH-related mHealth interventions around the world to investigate their appropriateness.

Methods

Search Strategies and Selection Criteria

We followed the methods detailed in a peer-reviewed systematic review protocol that is registered with International Prospective Register of Systematic Reviews (PROSPERO; CRD42017055570).

Three relevant English databases (PubMed, EMBASE, and Global Health) and 3 major Chinese databases (China National Knowledge Infrastructure, VIP Database for Chinese Technical Periodicals, and Wanfang Data Knowledge Service Medium) were searched. The search terms comprised key words from the following 3 dimensions: mHealth, maternal health care, and child health care. The searching strategy for each database was developed on the basis of key words identified from the literature and rules of subject headings in each database. Multimedia Appendix 1 shows details of the search terms used.

There has been significant development of mHealth [8] in medicine. This research paper examined studies in English and Chinese published between 1 January 2011 and 31 December 2016 (due to the language ability of the researchers). In addition, the references of included papers were reviewed to identify relevant papers.

Studies were eligible for inclusion if they aimed at improving RMNCH and studied interventions were conducted through mobile phone or tablet. The researchers focused on children under 6 years, as they are the most vulnerable group and a major target of mHealth interventions. Studies were excluded if their interventions were phone call alone, or functions and implementation were not clearly described, or they were descriptions of information technology. Systematic reviews or commentaries were also excluded, but kept as references. To achieve high sensitivity on search terms (but the accuracy was relatively low), no restriction was placed on study design and disease type. The corresponding authors were contacted if descriptions of interventions or studies were not clear enough for inclusion or exclusion.

All the searches were conducted on 21 February 2017. Searches were done independently by 2 reviewers, and a supervisor was invited for independent arbitration where consensus was not reached.

Data Extraction

The following types of data were extracted: (1) basic information of research, such as the author, publication year, and the country and region where the interventions were delivered; (2) the target population, health care stages, and the health issue corresponding to a particular intervention; (3) the type of mHealth medium (App or short message service [SMS]), and the description and function of the health intervention; (4) the study design and the number of participants given the health interventions; and (5) the primary outcome and the results of the study. The data extraction and quality assessment were processed by 2 independent reviewers, and any disagreement was resolved by a supervisor. The categorization of functions was based on mHealth and ICT Framework for mHealth innovations in the RMNCH field, which had 12 common mHealth applications used as health system strengthening innovations [11].

Data Analysis

The studies were not limited to randomized controlled trials (RCTs), as the researchers aimed to analyze and present the characteristics of RMNCH-related mHealth interventions described in all the reviewed studies. The descriptive analysis of the main characteristics and the key findings were processed and presented. For RCTs, we presented their results on effectiveness as categorical variables. Studies that showed

significantly better effectiveness in the intervention group compared with the control group were recorded as positive or recorded as negative or not clear if no statistical data were presented.

Substantial heterogeneity existed among the studies. As a result, the researchers were only able to perform random-effects meta-analysis using the inverse variance method for the 5 comparable studies, 2 studies on exclusive breastfeeding (EBF) and 3 studies on antenatal check (ANC). Meta-analyses were undertaken and the bias of the 5 studies was evaluated using Review Manager 5.3 (Cochrane Collaboration).

Results

From 6 databases, 5140 papers were identified in the initial search for screening. Finally, 245 published papers (studies) were included in this systematic review, among which 20.8% (51/245) studies were RCTs and 24.9% (61/245) were quasi-experimental studies. Two studies [17,18] on EBF and three studies [19-21] on ANC were included in the meta-analysis. The search process is shown in Figure 1.

Figure 1. Identification process for eligible studies.

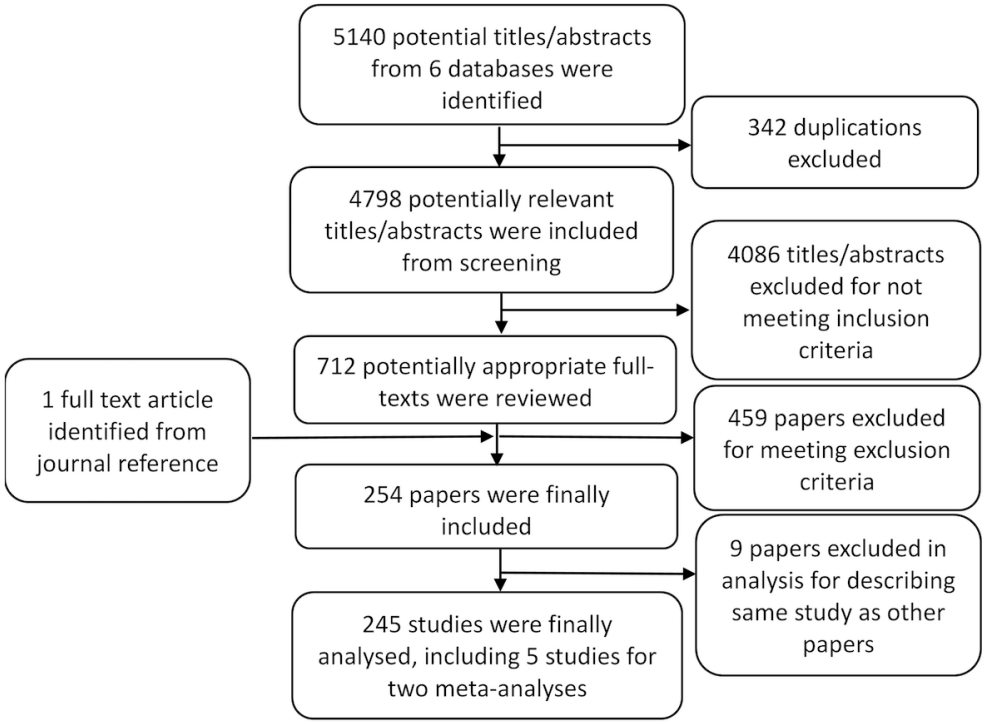


Figure 2. Trend of studies published from 2011 to 2016. Interventions combined with SMS and App were counted twice in each medium.

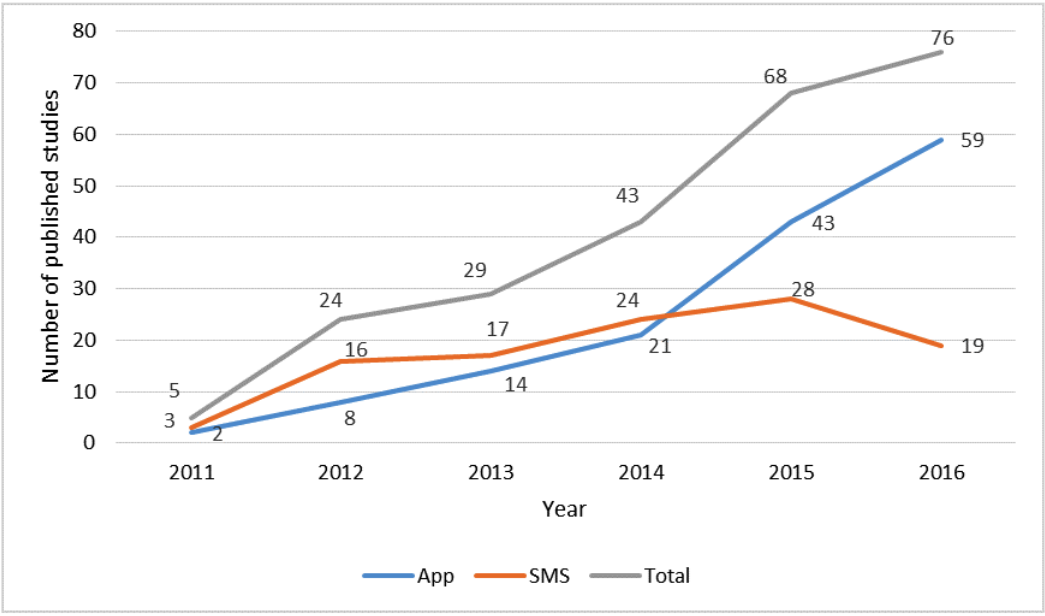


Figure 2 shows that the number of publications on mHealth in RMNCH increased significantly from 2011 to 2016. Among the 245 studies, interventions in 40% (98/245) studies were carried out through the SMS medium, 56.3% (138/245) through mobile or tablet-based apps (including 2 Web-based Apps), and 3.8% (9/245) combining SMS and App concurrently. The number of studies on the App also saw a steady increase from 2011 to 2016, whereas the number of studies on SMS increased slightly from 2011 to 2015 and then decreased from 2015 to 2016.

The findings and characteristics of the studies are presented in Table 1, including geographic distribution, medium, target population, health care stage, health issue addressed, and study design. The 245 studies came from 46 countries and 6 regions, and studies from the West Pacific, the Americas, and Africa regions accounted for 78.3% of all studies.

The targeted populations of mHealth interventions included both health service beneficiaries and providers. However, the number of studies targeting health service beneficiaries (n=176) was 3 times higher than that of health service providers (n=55). Only a few (n=14) considered both beneficiaries and providers.

This study divided RMNCH care into 5 stages, namely, prepregnancy, pregnancy, delivery, postpartum, and childcare. Although 76.3% (187/245) of the studies investigated mHealth interventions for childcare and pregnancy, only 6.5% (16/245) studies paid attention to delivery and postpartum stages. Ninety-one percent of mHealth interventions (224/245) focused on single stage, whereas 8.5% (21/245) of interventions designed for 2 and/or more stages at the same time.

The functions of mHealth interventions were categorized into 16 types (Table 2), which largely overlap with the ICT Framework, except the functions that optimize hospital service flow, such as setting an appointment (to set doctor's appointments with hospital) and laboratory results (to check laboratory results through an App linked with the hospitals' information system). Meanwhile, functions relating to human resource management and supply chain management described by the ICT framework were not identified in this study. Health education or promotion (to provide users with health information and lifestyle advice), physical or bio data monitoring (to monitor physical or bio data of patients in distance and to adjust for treatment in real time, especially for chronic conditions), and reminders (to remind users for antenatal checks, the ovulation time, medication, etc) are the most commonly used functions. Apart from health education or promotion, there is a difference in the most frequently adopted functions among App and SMS interventions. That is, the reminders (27.1%) and data collection and management (to collect data for research or administrative purposes) (15.9%) ranked second and third most popular functions in SMS interventions, respectively, whereas physical or bio data monitoring (36.1%) and counseling (to consult a health professional directly) (17.7%) ranked second and third most popular in App interventions, respectively. mHealth interventions with complex algorithms were more frequently observed in Apps compared with SMSs, such as decision support and diagnosis. The number of mHealth interventions with single functions was more than twice of that with multi-functions (≥2 functions), which were more frequently seen in the App.

Studies on SMS and/or App interventions were distributed unevenly among regions. In Africa, the number of SMS-based studies (65.5%) was about 3 times higher than that of App-based studies (32.7%), whereas App-based studies played a dominant role in South-East Asia (80.0%), Europe (75.0%), the Eastern Mediterranean (66.7%), and the Americas (59.7%). Meanwhile, the most frequently adopted functions in each region were associated with the type of mediums (SMS and/or App) dominantly applied in that region.

For the 51 identified RCTs, we presented their results in Table 3 with respect to whether the results showed that mHealth interventions were significantly effective. More than half (n=29) of the studies had positive results supporting the effectiveness of mHealth interventions and 43.1% (n=22) had negative or unclear results. Details about the effectiveness of mHealth interventions based on different mediums, functions, health issues, and the stages are shown in Table 3.

Two studies (Flax et al [17] and Jiang et al [18]) compared the effect of mHealth interventions using SMS compared with routine health care, encouraging breastfeeding in Nigeria and China. The results of both trials showed that the rates of EBF for 6 months were higher in the intervention group than in the control group. We undertook meta-analysis of the effect of mHealth intervention versus routine health care on EBF for 6 months. The merged estimates showed that the rate of EBF for 6 months was higher in the mHealth intervention groups compared with the control group (OR 2.03, 95% CI 1.34-3.08, I^2=25%; Figure 3).

Lund et al [19], Luo et al [20], and Shiferaw et al [21] compared the effect of mHealth interventions versus routine care on ANC in Zanzibar (SMS), China (SMS), and Ethiopia (App), respectively. The results of all trials showed that the rates of 4 or more ANCs were higher in the mHealth group than in the control group. The merged estimates from the meta-analysis showed that the rates of 4 or more ANCs (OR 1.43, 95% CI 1.13-1.79, I^2=78%; Figure 4) were higher in the mHealth intervention groups than in the control groups.

Table 1. Characteristics of included studies.

Category	Studies, n (%)
Region	
The Americas	67 (27.3)
Europe	32 (13.1)
The Western Pacific	70 (28.6)
South-East Asia	15 (6.1)
The Eastern Mediterranean	6 (2.4)
Africa	55 (22.4)
Health issues[a]	
Infectious diseases	28 (11.4)
Chronic diseases	43 (17.6)
Mental and behavioral disorders	11 (4.5)
Essential RMNCH[b] issues	16 (66.5)
Study design[c]	
mHealth product description	87 (35.5)
Quasi-experiment	61 (24.9)
RCT[d]	51 (20.8)
Cross-sectional study	21 (8.6)
RCT protocol	19 (7.7)
Qualitative study	5 (2.0)
Case report	1 (0.4)
Medium	
SMS[e]	98 (40.0)
App	138 (56.3)
SMS and App	9 (3.8)
Target population	
Health service beneficiaries	176 (71.8)
Women[f]	94 (38.4)
Parents	76 (31.0)
Children	11 (4.5)
Health service providers	55 (22.4)
Health professionals	29 (11.8)
Health workers and volunteers	24 (9.8)
Administrators	3 (1.2)
Beneficiaries and providers	14 (5.7)
Health care stages	
Prepregnancy	21 (8.6)
Pregnancy	68 (27.8)
Delivery and postpartum	16 (6.5)
Childcare	119 (48.6)
Multi-stages	21 (8.6)

[a]Forty-seven types of health issues were identified from included studies and divided into 4 categories based on the 10[th]version of the International

Statistical Classification of Diseases and Related Health Problems (International Classification of Diseases-10) and data availability. The details on these health issues can be found in Multimedia Appendix 2.

[b]RMNCH: Reproductive, maternal, newborn, and child health.

[c]Details on study design can be found in Multimedia Appendix 3.

[d]RCT: randomized controlled trial.

[e]SMS: short message service.

[f]Women including women at child-bearing age, pregnant women, and perinatal women.

Table 2. Functions of mHealth interventions delivered by short message service and App.

Function[a]	Total, N (%)	SMS[b], n (%)	App, n (%)
Single function	162 (66.1)	75 (46.3)	92 (56.8)
Two functions	69 (28.2)	30 (43.5)	42 (60.9)
Three and more functions	14 (5.7)	2 (14.3)	13 (92.9)
Health education or promotion	110 (44.9)	60 (56.1)	53 (36.1)
Physical or bio data monitoring	58 (23.7)	10 (9.3)	53 (36.1)
Reminders	40 (16.3)	29 (27.1)	12 (8.2)
Counseling	38 (15.5)	15 (14.0)	26 (17.7)
Data collection and management	37 (15.1)	17 (15.9)	20 (13.6)
Decision support and guideline	18 (7.3)	0 (0.0)	18 (12.2)
Diagnosis and treatment	17 (6.9)	2 (1.9)	17 (11.6)
Appointment making	11 (4.5)	4 (3.7)	7 (4.8)
On-the-job training for health professionals	9 (3.7)	2 (1.9)	7 (4.8)
Laboratory results	5 (2.0)	0 (0.0)	5 (3.4)
Communication	4 (1.6)	1 (0.9)	3 (2.0)
Payment	3 (1.2)	0 (0.0)	3 (2.0)
Supervision and technical support	2 (0.8)	0 (0.0)	2 (1.4)
Hospital guidelines	2 (0.8)	0 (0.0)	2 (1.4)
Cash transfer	1 (0.4)	1 (0.9)	0 (0.0)
Electronic health record check	1 (0.4)	0 (0.0)	1 (0.7)

[a]Studies with multiple functions were counted repeatedly in each function category.

[b]SMS: short message service.

Table 3. mHealth interventions and results from randomized clinical trials.

Category	Total studies, N	RCTs[a], n	RCTs with positive results, n (%)
Total	245	51	29 (56.9)
Medium			
App	147	13	9 (69.2)
SMS[b]	107	40	22 (55.0)
Function			
Studies with multiple functions[c]	83	13	9 (69.2)
Studies with single functions	162	38	20 (52.6)
Health issue			
Essential RMNCH[d] issues	163	37	21 (56.8)
Other RMNCH-related diseases[e]	82	14	8 (57.1)
Stages			
Prepregnancy	26	10	6 (60.0)
Pregnancy	88	14	11 (78.6)
Delivery and postpartum	28	4	3 (75.0)
Child care	139	24	10 (41.7)

[a]RCTs: randomized controlled trials.

[b]SMS: short message service.

[c]Multiple functions refer to 2 or more functions, such as health education or promotion, physical or biodata monitoring, and reminders concurrently.

[d]RMNCH: reproductive, maternal, newborn, and child health.

[e]Other RMNCH-related diseases include infectious diseases, chronic diseases, and mental and behavioral disorders.

Figure 3. Meta-analysis of the effect of mHealth intervention versus routine prenatal care on exclusive breastfeeding for 6 months based on two studies undertaken in Nigeria and China.

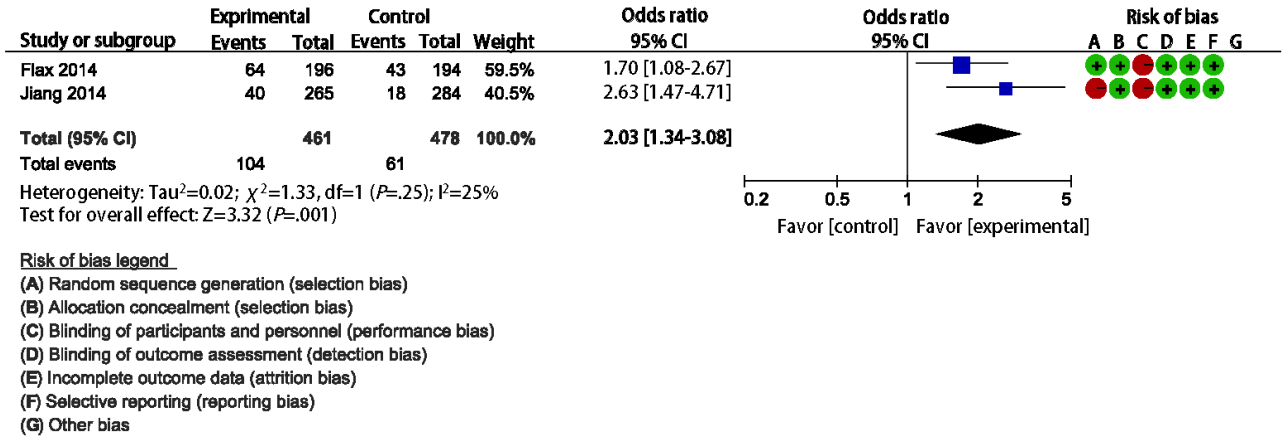
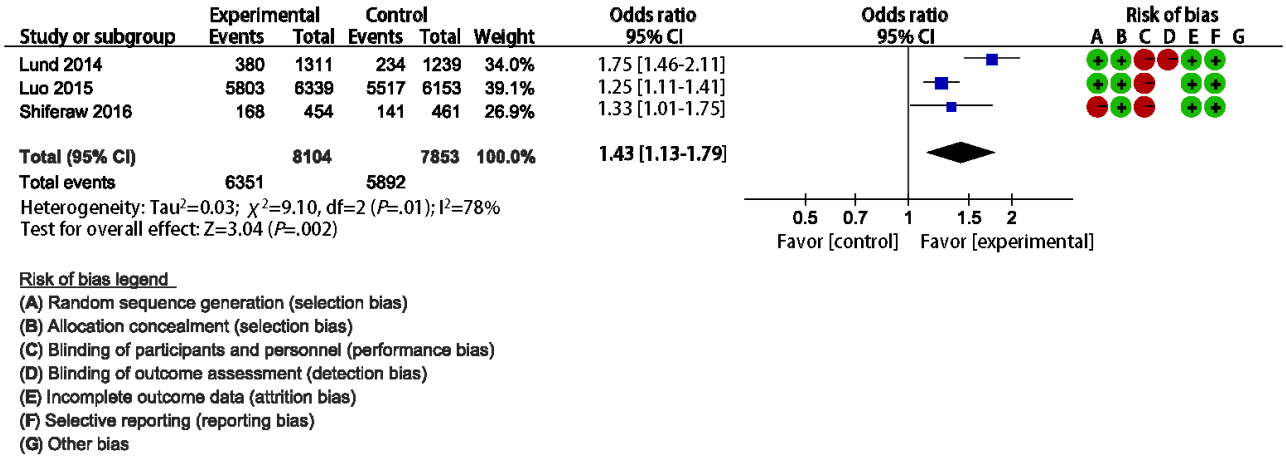


Figure 4. Meta-analysis of the effect of mHealth intervention versus routine prenatal care on four or more antenatal check rates based on three studies undertaken in Zanzibar, China, and Ethiopia.



Discussion

Principal Findings

The study found a rapid increase in the number of publications on mHealth interventions for RMNCH, especially for those using App. However, the overall number of publications included (n=245) remained relatively small, among which only 51 RCTs were identified. Although 2 meta-analyses based on 2 studies on EBF and 3 studies on ANC showed positive evidences to support the effectiveness of mHealth interventions, almost half (43.2%) of RCTs showed negative or unclear results of mHealth interventions. In addition, the studies were too heterogeneous and dispersed to generate merged results for individual mHealth interventions on specific health issues. Among all interventions identified in this study, the ones designed for health care beneficiaries exceeded 3 times those for health care providers. mHealth interventions with a single function or stage were dominant among all identified studies.

mHealth interventions are popular worldwide with fast upgrades of technologies and improvement of communication infrastructure. However, the overall number of relevant publications remains relatively small compared with other interventions for RMNCH [22]. Consistent with Sahu's finding [23], Apps seem to dominate the mHealth market compared with SMS, as they can provide multiple and complex functions simultaneously.

More than one-fifth of the studies (22.4%) were from Africa, which is comparable to that from Western Pacific (28.6%) and the Americas (27.3%). This can be partially explained by the rapid development of communication infrastructure and high ownership rate of mobile phones in Africa in recent years [24,25]. Tailored messages and other mHealth interventions through mobile phone can reach target populations and the hard-to-reach population more efficiently compared with traditional approaches.

Evidences on the effectiveness of single or multiple mHealth functions on individual or multiple health conditions were far from adequate. A high proportion of descriptive and observational studies and small number of high-quality RCTs identified indicate that evaluation of mHealth innovation designed for RMNCH is in its infancy. This led to a result that diversified outcome indicators and health issues were reported from a very limited number of RCTs, which placed huge challenges in acquiring merged effects of mHealth interventions.

Although the results of 2 meta-analyses on EBF [17,18] and ANC [19-21] supported the use of mHealth interventions, almost half of the RCTs still showed negative or inconclusive results of mHealth interventions, which could be explained in 3 ways. First, most RCTs with negative or inconclusive results have a small sample size (less than 200 people) or a short follow-up time (less than 6 months). Second, major outcome measurements, such as blood glucose and weight, are hard to be improved in a short follow-up period. Third, most RCTs were less rigorous in study design, as the evaluation of mHealth interventions is still at an early stage. Therefore, the results of these included studies need to be further tested to reduce the risk of bias derived from any inherent weakness in the study designs.

Although the number of high-quality RCTs identified for the health service provider–related functions was scarce, many quasi-experimental studies [26-29] provided some preliminary evidence in favor of functions, including decision support (to provide clinical guidelines for health professionals when necessary) and on-the-job training (to conduct trainings for health professionals to improve service quality). This may shed some light for future study, as providing technical support to health service provider is crucial in improving health care quality and retention of service providers from health system strengthening perspective.

Interventions delivered through SMS could only enable functions with simple algorithms, such as health education and reminders. One of our previous studies found that almost all existing RMNCH Apps in the Chinese market were embedded with multiple functions and covered multiple stages through a person's life course. However, most of the studies included in this review focused on single function and stage, which reveals a gap between research and practice.

Apart from commonly concerned ANC and general childcare (such as feeding and immunization), studies from developing countries largely focused on infectious disease and essential maternal and childcare (such as human immunodeficiency virus [HIV] and diarrhea), whereas studies from developed countries explored more on noncommunicable diseases (such as asthma, gestational diabetes mellitus, and cancer). However, disease prevalence is changing and burden of NCDs is becoming devastating in those developing countries, according to WHO's latest report on global disease burden [30,31]. Therefore, NCDs should be given more consideration when designing mHealth interventions and studies in developing countries.

This study also found that SMS was used more frequently in low-middle income countries compared with Apps, providing basic functions such as health education or promotion [32,33], reminder [34-36], and data collection [37,38]. RCTs on SMS showed some positive results in the prevention of mother-to-child transmission (PMTCT), childhood disease management, and family planning. However, the evidence is not adequate to draw conclusions that one function is more effective for a particular health condition when delivered through either SMS or App. When it comes to the selection of SMS over App, cost-effectiveness analysis need to be considered, as the cost for SMS-based functions usually is much cheaper than that for App-based functions [39].

Conclusions

The major limitation of this study is that only 6 databases were searched. This can result in missing of high-quality RCTs on mHealth intervention for RMNCH, which may contribute to merged effect and lend more weight to the effectiveness of some mHealth interventions. In summary, published studies on RMNCH-related mHealth interventions are increasing, but have been far from adequate in evaluating the effectiveness of such interventions on individual health issues. More rigorous evaluations are needed to draw consistent conclusions. The studied mHealth interventions were relatively simple. More research is needed to evaluate mHealth products with multiple functions or stages, especially those popular outside clinical practice.

Acknowledgments

This study was funded by Qualcomm Wireless Reach and China Children and Teenagers' Fund. The sponsors had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author (Puhong Zhang) had the final responsibility for decision for publication.

Authors' Contributions

PZ and HC designed the study, and HC designed search strategies. HC and YC conducted the literature search, screening, and data extraction. YC and HC conducted data analysis and contributed equally to data interpretation and writing. PZ contributed to data interpretation and manuscript revision. LD contributed to data collection, and WN contributed to data interpretation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies.

[PDF File (Adobe PDF File), 212KB - mhealth_v6i1e7_app1.pdf]

Multimedia Appendix 2

Details of health issues targeted by mHealth interventions.

[PDF File (Adobe PDF File), 103KB - mhealth_v6i1e7_app2.pdf]

Multimedia Appendix 3

Study designs.

[PDF File (Adobe PDF File), 260KB - mhealth_v6i1e7_app3.pdf]

References

1. World Bank. Switzerland: World Health Organization; 2015. Trends in maternal mortality: 1990-2015: estimates from WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division URL: http://documents.worldbank.org/curated/en/798471467987911213/pdf/100949-v2-P152335-PUBLIC-Box394828B-Trends-in-Maternal-Mortality-1990-to-2015-full-report.pdf [WebCite Cache ID 6vRuemX2m]

2. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. Lancet 2015 Dec 05;386(10010):2275-2286. [doi: 10.1016/S0140-6736(15)00120-8] [Medline: 26361942]
3. Waage J, Banerji R, Campbell O, Chirwa E, Collender G, Dieltiens V, et al. The millennium development goals: a cross-sectoral analysis and principles for goal setting after 2015 Lancet and London international development centre commission. Lancet 2010 Sep 18;376(9745):991-1023. [doi: 10.1016/S0140-6736(10)61196-8] [Medline: 20833426]
4. Sustainabledevelopment. New York: The United Nations; 2015. Sustainable development goals: the United Nations URL: https://sustainabledevelopment.un.org/?menu=1300 [WebCite Cache ID 6vRukyK37]
5. Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Child Health Epidemiology Reference Group of WHO and UNICEF. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 2012 Jun 09;379(9832):2151-2161 [FREE Full text] [doi: 10.1016/S0140-6736(12)60560-1] [Medline: 22579125]
6. Malvey D, Slovensky D. MHealth: Transforming Healthcare. New York: Springer; 2014.
7. International Telecommunication Union. The world in 2014: ICT facts and figures. Geneva, Switzerland: International Telecommunication Union; 2014. URL: http://wwwa.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf [WebCite Cache ID 6vRv7q3Z4]
8. World Health Organization. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. Switzerland: World Health Organization; 2011. URL: http://www.who.int/ehealth/mhealth_summit.pdf [WebCite Cache ID 6vRvdiC1k]
9. Agarwal S, Labrique A. Newborn health on the line: the potential mHealth applications. JAMA 2014 Jul 16;312(3):229-230. [doi: 10.1001/jama.2014.6371] [Medline: 24953141]
10. Qiang C, Yamamichi M, Hausman V, Altman D, Unit I. World Bank. Washington: World Bank; 2011. Mobile applications for the health sector URL: http://documents.worldbank.org/curated/en/751411468157784302/pdf/726040WP0Box370th0report00Apr020120.pdf [accessed 2017-12-03] [WebCite Cache ID 6vRvl6u03]
11. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
12. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
13. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. Effectiveness of mHealth interventions for maternal, newborn and child health in low- and middle-income countries: Systematic review and meta-analysis. J Glob Health 2016 Jun;6(1):010401 [FREE Full text] [doi: 10.7189/jogh.06.010401] [Medline: 26649177]
14. Sondaal SFV, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
15. Al Dahdah M, Desgrees Du Lo UA, Meadel C. Mobile health and maternal care: a winning combination for healthcare in the developing world. Health Policy Technol 2015 Apr;4(3):225-231 [FREE Full text]
16. Archangeli C, Marti FA, Wobga-Pasiah EA, Zima B. Mobile health interventions for psychiatric conditions in children: a scoping review. Child Adolesc Psychiatr Clin N Am 2017 Jan;26(1):13-31. [doi: 10.1016/j.chc.2016.07.009] [Medline: 27837939]
17. Flax VL, Negerie M, Ibrahim AU, Leatherman S, Daza EJ, Bentley ME. Integrating group counseling, cell phone messaging, and participant-generated songs and dramas into a microcredit program increases Nigerian women's adherence to international breastfeeding recommendations. J Nutr 2014 Jul;144(7):1120-1124 [FREE Full text] [doi: 10.3945/jn.113.190124] [Medline: 24812071]
18. Jiang H, Li M, Wen LM, Hu Q, Yang D, He G, et al. Effect of short message service on infant feeding practice: findings from a community-based study in Shanghai, China. JAMA Pediatr 2014 May;168(5):471-478 [FREE Full text] [doi: 10.1001/jamapediatrics.2014.58] [Medline: 24639004]
19. Lund S, Nielsen BB, Hemed M, Boas IM, Said A, Said K, et al. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. BMC Pregnancy Childbirth 2014 Jan 17;14:29 [FREE Full text] [doi: 10.1186/1471-2393-14-29] [Medline: 24438517]
20. Luo M. Effectiveness of health education intervention through mobile telephone message in pregnancy. Changsha, China: Central South University; 2013.
21. Shiferaw S, Spigt M, Tekie M, Abdullah M, Fantahun M, Dinant G. The effects of a locally developed mHealth intervention on delivery and postnatal care utilization; a prospective controlled evaluation among health centres in Ethiopia. PLoS One 2016;11(7):e0158600 [FREE Full text] [doi: 10.1371/journal.pone.0158600] [Medline: 27383186]
22. Yao N, Wang F, Wang X, Yu Q, Liu X. Research in mother and child health organization of China: a bibliometric analysis. Chinese Journal of Medical Library and Information Science 2012(08):18-22.

23. Sahu M, Grover A, Joshi A. Role of mobile phone technology in health education in Asian and African countries: a systematic review. Int J Electron Healthc 2014;7(4):269-286. [doi: 10.1504/IJEH.2014.064327] [Medline: 25161104]
24. Aker JC, Mbiti IM. Mobile phones and economic development in Africa. J Econ Perspect 2010;24(3):207-232 [FREE Full text] [doi: 10.1257/jep.24.3.207]
25. Traore A. Leo-univ-orleans. Orleans, France: University of Orleans; 2012. Mobile phone in Sub-Saharan Africa: impact on technological innovation and economic development URL: http://data.leo-univ-orleans.fr/media/seminars/235/traore-awa.pdf [accessed 2017-12-07] [WebCite Cache ID 6vXKh3o8M]
26. Hamilton M, Bora G. Evidence-based scale-up of mSakhi community health worker mHealth system in Uttar Pradesh, India. Annals of Global Health 2015;81(1):31 [FREE Full text]
27. Thukral A, Joshi M, Joshi P, Prakash V, Adkoli BV, Deorari AK. Apps for management of sick newborn: evaluation of impact on health care professionals. J Trop Pediatr 2014 Oct;60(5):370-376. [doi: 10.1093/tropej/fmu032] [Medline: 24924579]
28. McNabb M, Chukwu E, Ojo O, Shekhar N, Gill CJ, Salami H, et al. Assessment of the quality of antenatal care services provided by health workers using a mobile phone decision support application in northern Nigeria: a pre/post-intervention study. PLoS One 2015;10(5):e0123940 [FREE Full text] [doi: 10.1371/journal.pone.0123940] [Medline: 25942018]
29. Prakash V, Thukral A, Sankar MJ, Agarwal RK, Paul VK, Deorari AK. Efficacy and acceptability of an "App on sick newborn care" in physicians from newborn units. BMC Med Educ 2016 Mar 08;16:84 [FREE Full text] [doi: 10.1186/s12909-016-0579-3] [Medline: 26956397]
30. World Health Organization. Global status report on noncommunicable diseases 2014. Switzerland: World Health Organization; 2014. URL: http://www.wsmi.org/wp-content/uploads/2015/01/Global-Status-Report-NCDs-2014.pdf [WebCite Cache ID 6vS0kTpp5]
31. WHO. Switzerland: World Health Organization; 2017. Noncommunicable diseases: the slow motion disaster URL: http://www.who.int/publications/10-year-review/ncd/en/ [accessed 2017-12-04] [WebCite Cache ID 6vS1Dl1ih]
32. Grabosch S, Gavard JA, Mostello D. Text4baby improves glycemic control in pregnant women with diabetes. Am J Obstet Gynecol 2014(1):S88. [doi: 10.1016/j.ajog.2013.10.184]
33. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
34. Heatley E, Middleton P, Hague W, Crowther C. The DIAMIND study: postpartum SMS reminders to women who have had gestational diabetes mellitus to test for type 2 diabetes: a randomised controlled trial - study protocol. BMC Pregnancy Childbirth 2013 Apr 12;13:92 [FREE Full text] [doi: 10.1186/1471-2393-13-92] [Medline: 23587090]
35. Ahlers-Schmidt CR, Chesser AK, Nguyen T, Brannon J, Hart TA, Williams KS, et al. Feasibility of a randomized controlled trial to evaluate text reminders for immunization compliance in kids (TRICKs). Vaccine 2012 Aug 3;30(36):5305-5309. [doi: 10.1016/j.vaccine.2012.06.058] [Medline: 22750044]
36. Bigna JJR, Noubiap JJN, Kouanfack C, Plottel CS, Koulla-Shiro S. Effect of mobile phone reminders on follow-up medical care of children exposed to or infected with HIV in Cameroon (MORE CARE): a multicentre, single-blind, factorial, randomised controlled trial. Lancet Infect Dis 2014 Jul;14(7):600-608. [doi: 10.1016/S1473-3099(14)70741-8] [Medline: 24932893]
37. Odeny TA, Bukusi EA, Cohen CR, Yuhas K, Camlin CS, McClelland RS. Texting improves testing: a randomized trial of two-way SMS to increase postpartum prevention of mother-to-child transmission retention and infant HIV testing. AIDS 2014 Sep 24;28(15):2307-2312 [FREE Full text] [doi: 10.1097/QAD.0000000000000409] [Medline: 25313586]
38. Regan AK, Blyth CC, Effler PV. Using SMS technology to verify the safety of seasonal trivalent influenza vaccine for pregnant women in real time. Med J Aust 2013 Dec 16;199(11):744-746. [Medline: 24329634]
39. Larsen-Cooper E, Bancroft E, Rajagopal S, O'Toole M, Levin A. Scale matters: a cost-outcome analysis of an m-health intervention in Malawi. Telemed J E Health 2016 Apr;22(4):317-324 [FREE Full text] [doi: 10.1089/tmj.2015.0060] [Medline: 26348994]

Abbreviations

ANC: antenatal checks
EBF: exclusive breastfeeding
HIV: human immunodeficiency virus
ICT: information and communication technology
mHealth: mobile health
NCDs: noncommunicable diseases
OR: odds ratio
PMTCT: revention of mother-to-child transmission
RCT: randomized controlled trial
RMNCH: reproductive, maternal, newborn, and child health
SMS: short message service

WHO: World Health Organization

Edited by G Eysenbach; submitted 20.09.17; peer-reviewed by M AlDahdaha, R Pankomera; comments to author 11.10.17; revised version received 23.10.17; accepted 29.10.17; published 09.01.18.

Please cite as:
Chen H, Chai Y, Dong L, Niu W, Zhang P
Effectiveness and Appropriateness of mHealth Interventions for Maternal and Child Health: Systematic Review
JMIR Mhealth Uhealth 2018;6(1):e7
URL: http://mhealth.jmir.org/2018/1/e7/
doi:10.2196/mhealth.8998
PMID:29317380

Review

The Impact of mHealth Interventions: Systematic Review of Systematic Reviews

Milena Soriano Marcolino[1], MD, MSc, PhD; João Antonio Queiroz Oliveira[1], PharmD, MSc; Marcelo D'Agostino[2], MSc; Antonio Luiz Ribeiro[1], MD, PhD; Maria Beatriz Moreira Alkmim[1], MD, MSc; David Novillo-Ortiz[2], MLIS, MSc, PhD

[1]Medical School and Telehealth Center, University Hospital, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[2]Pan American Health Organization, Washington, DC, United States

Corresponding Author:
David Novillo-Ortiz, MLIS, MSc, PhD
Pan American Health Organization
525 23rd St NW
Washington, DC, 20037
United States
Phone: 1 2028124726
Email: novillod@paho.org

Abstract

Background: Mobile phone usage has been rapidly increasing worldwide. mHealth could efficiently deliver high-quality health care, but the evidence supporting its current effectiveness is still mixed.

Objective: We performed a systematic review of systematic reviews to assess the impact or effectiveness of mobile health (mHealth) interventions in different health conditions and in the processes of health care service delivery.

Methods: We used a common search strategy of five major scientific databases, restricting the search by publication date, language, and parameters in methodology and content. Methodological quality was evaluated using the Measurement Tool to Assess Systematic Reviews (AMSTAR) checklist.

Results: The searches resulted in a total of 10,689 articles. Of these, 23 systematic reviews (371 studies; more than 79,665 patients) were included. Seventeen reviews included studies performed in low- and middle-income countries. The studies used diverse mHealth interventions, most frequently text messaging (short message service, SMS) applied to different purposes (reminder, alert, education, motivation, prevention). Ten reviews were rated as low quality (AMSTAR score 0-4), seven were rated as moderate quality (AMSTAR score 5-8), and six were categorized as high quality (AMSTAR score 9-11). A beneficial impact of mHealth was observed in chronic disease management, showing improvement in symptoms and peak flow variability in asthma patients, reducing hospitalizations and improving forced expiratory volume in 1 second; improving chronic pulmonary diseases symptoms; improving heart failure symptoms, reducing deaths and hospitalization; improving glycemic control in diabetes patients; improving blood pressure in hypertensive patients; and reducing weight in overweight and obese patients. Studies also showed a positive impact of SMS reminders in improving attendance rates, with a similar impact to phone call reminders at reduced cost, and improved adherence to tuberculosis and human immunodeficiency virus therapy in some scenarios, with evidence of decrease of viral load.

Conclusions: Although mHealth is growing in popularity, the evidence for efficacy is still limited. In general, the methodological quality of the studies included in the systematic reviews is low. For some fields, its impact is not evident, the results are mixed, or no long-term studies exist. Exceptions include the moderate quality evidence of improvement in asthma patients, attendance rates, and increased smoking abstinence rates. Most studies were performed in high-income countries, implying that mHealth is still at an early stage of development in low-income countries.

(JMIR Mhealth Uhealth 2018;6(1):e23) doi:10.2196/mhealth.8873

KEYWORDS

telemedicine; medical informatics; mobile phones

Introduction

Mobile phone usage has been rapidly increasing worldwide [1,2]. In many high-income countries, mobile phone subscriptions exceed the population, and in many low- and middle-income countries, this number is expanding faster than other infrastructures [2]. Mobile technology's mobility, instantaneous access, and direct communication allow for faster transfer of health information, which in turn supports medical and public health practices. These characteristics define mobile health (mHealth). mHealth could transform the worldwide delivery of health services, especially in low- and middle-income countries. This includes simple apps and complex technologies including voice, text messaging (short message service, SMS), multimedia message service, Bluetooth technology, and others [3].

mHealth is increasingly being used (1) for patient communication, monitoring, and education, (2) to reduce the burden of diseases linked with poverty, (3) to improve access to health services, clinical diagnosis, and treatment adherence, and (4) for chronic disease management [4-6].

It is commonly stated that mHealth effectively improves the quality of care and that it can quickly be adapted on a large scale and at low cost, but evidence regarding its effectiveness and cost-effectiveness is still lacking in different areas. As the evidence in this field is consistently growing, many systematic reviews have already been performed. A thorough review of available evidence is essential to guide clinical and health policy decisions. Consequently, complex reviews, which may assess multiple interventions, different or distinct populations, and different outcomes may adequately support health policy decision making in this context [7]. Therefore, the objective of this study was to perform a systematic review of systematic reviews that assessed the effectiveness of mHealth interventions in different health conditions and in the processes of health care service delivery, in order to investigate for which areas there is evidence and which areas still require further studies.

Methods

This study is a systematic review of systematic reviews and is part of a series of four reviews that assessed the impact to telehealth strategies in different health conditions and in health care service delivery. The study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and the methodological considerations when using existing systematic reviews [7].

Search Strategy and Inclusion Criteria

A literature search was performed using MEDLINE (accessed by PubMed), IEEE Xplore Digital Library, Cochrane (Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Cochrane Methodology Register, Database of Abstracts of Reviews of Effects, National Health Service Economic Evaluation Database), *Literatura Latino-Americana e do Caribe em Ciências da Saúde* (LILACS), and *Indice Bibliográfico Español de Ciencias de la Salud* (IBECS) in November 2015. Cochrane, LILACS, and IBECS were assessed by Virtual Health Library (*Biblioteca Virtual em Saúde*). The search was restricted to studies in humans, publication date (from January 1, 2000, up to the search data), and publication language (English, French, Spanish, Italian, and Portuguese).

We used a common search strategy and allocated relevant studies to their respective reviews before assessing their risk of bias and extracting data. The search strategies for each database are given in Multimedia Appendix 1. All studies were included in the software StArt (State of the Art through Systematic Review) [8]. In this software, different combinations of the terms "systematic" and "review" identified systematic reviews by title or abstract.

An additional search using the same terms and parameters was performed in February 2016. The new search was more specific to make assessment manageable and was supplemented by a manual search of reference lists [9].

Study Selection

Systematic reviews covering the effectiveness or cost-benefit analysis of eHealth interventions were included. Exclusion criteria were (1) studies about feasibility, user acceptance, or usability, (2) studies that assessed "perceived benefits," and (3) nonsystematic reviews.

Initial screening was based on titles and abstracts, and articles were independently evaluated. Abstracts lacking information were retrieved for full-text evaluation. Subsequently, 2 investigators independently evaluated full-text articles and determined eligibility. Authorship, journal, or years were not blinded.

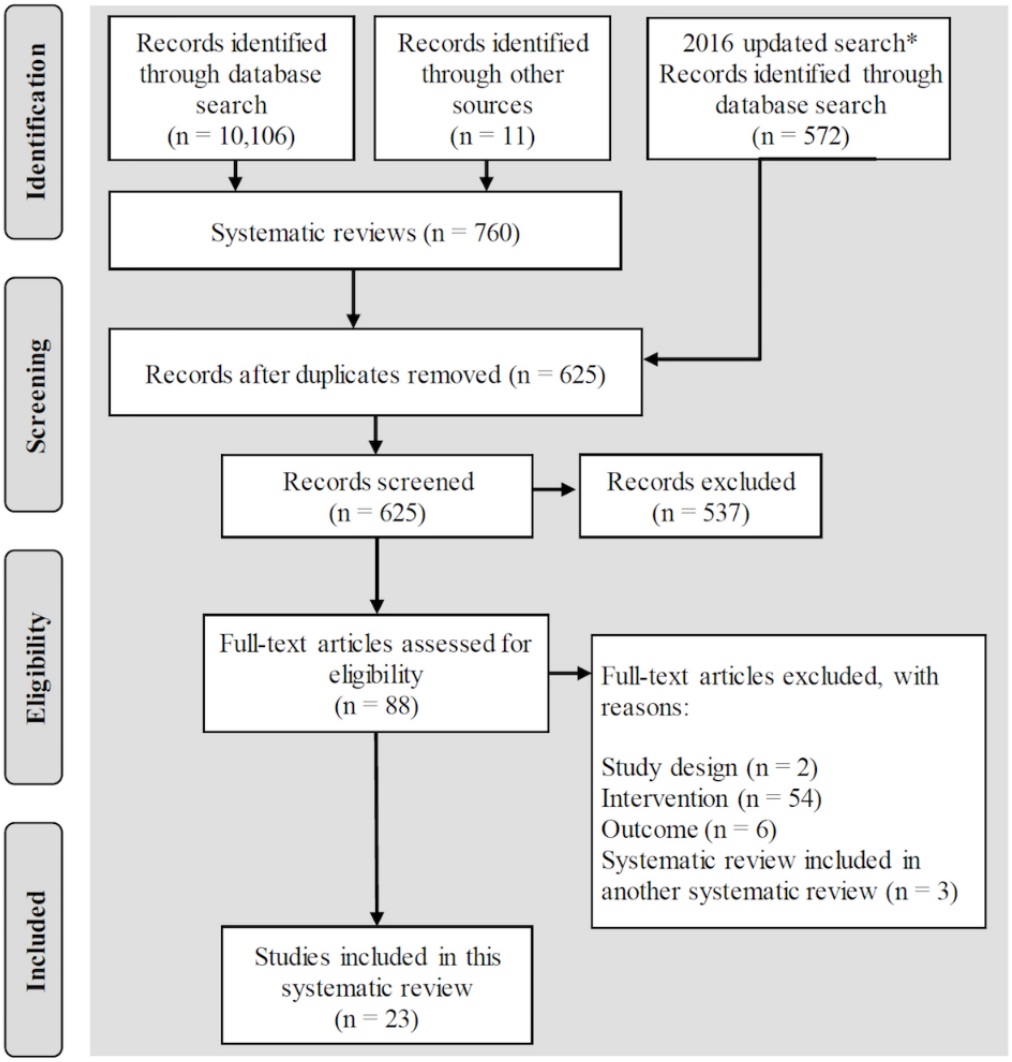
Data Extraction and Quality Assessment

Five investigators conducted data extraction following standardized criteria, and results were reviewed by 2 senior researchers. The following data were extracted: journal, publication year, databases searched, time period, setting/scenario, theme/specialty, objective, intervention type, number of studies, total number and countries of patients, study design, whether a review of systematic reviews or meta-analysis was performed, outcomes, main results, lessons and barriers for implementation, and main limitations. For cost analysis, the type and the perspective (ie, patient, health care provider, and/or society) were also extracted. Studies were evaluated using the Measurement Tool to Assess Systematic Reviews (AMSTAR) checklist for assessing methodological quality [10].

Results

A flow diagram of literature search and study selection results is shown in Figure 1. The database first search resulted in 10,106 articles, the updated search resulted in 572 articles, and 11 studies were found from additional sources. After exclusion of duplicates, 625 articles were screened and 537 were excluded. Full text of 88 eligible articles was reviewed. Out of these, 62 were excluded for not meeting the criteria relating to study type, intervention, or outcome. Three studies [5,11,12] were excluded for being included in a systematic review of systematic reviews that was included in this manuscript, to avoid duplication. The 23 studies remaining were included in this systematic review.

Figure 1. Flow of information through the different phases of the systematic review. Asterisk indicates that this search was limited to systematic reviews.



Descriptive Analysis of the Systematic Reviews

General Characteristics of Reviewed Papers

The 23 systematic reviews included (Multimedia Appendix 2) were published between 2009 and 2016 in 16 journals. The systematic reviews involved 371 studies. After verifying the sample size of each study, we found that at least 79,665 participants were included. In these reviews, systematic literature searches were performed from 1950 to April 2015 (see Multimedia Appendix 2).

Of the reviews, 17 included studies that were performed in low- and middle-income countries, 13 included studies in multiple settings, 6 specified particular settings, and 2 did not describe the setting. mHealth modalities described were mainly apps for chronic diseases, but also for disease management, treatment adherence, and changes in health behavior. Nine studies performed a meta-analysis [13-21].

Objective

The main objective of the reviews was to analyze the impact or effectiveness of mHealth interventions on chronic and noncommunicable diseases. Other focuses were to analyze mHealth in supporting chronic diseases management, health behavior change, attendance at appointments, disease and rehabilitation management, and the use of mHealth strategies by health workers.

Intervention

Different devices were used, including mobile phones, smartphones, personal digital assistants, MP3, phone plus app, medical device connected to phone by cord or wirelessly, and many others.

The most frequent intervention was SMS for reminders, education, motivation, or prevention. Sensors and point-of-care diagnosis, data collection, provider-provider communication, patient-provider communication, decision support, client education, provider work planning, training, protocol-based treatment, voicemail, videos, immediate physician feedback from a central location, cloud-based interactive voice response, disease management calls, disease monitoring, automated email to clinicians, treatment adherence, and phone counseling were also used.

These interventions were performed for smoking cessation, to increase physical activity, chronic disease management,

chemotherapy-related symptoms monitoring, sexual health behavior safety, alcohol consumption reduction, medication adherence, appointment attendance, stress management and anxiety reduction, vaccination timeliness, prenatal support, reduction in emergency referrals or adverse events, health information access, cardiopulmonary resuscitation skills, patient satisfaction, and social functioning.

Hall et al [22] categorized 12 common applications: (1) client education and behavior change, (2) sensors and point-of-care diagnostics, (3) registries and vital events tracking, (4) data collection and reporting, (5) electronic health records, (6) electronic decision support: information, protocols, algorithms, checklists, (7) provider-provider communication: user groups and consultation, (8) provider work planning, (9) provider training and education, (10) human resource management, (11) supply chain management, and (12) financial transactions and incentives.

Multiple interventions were used on significantly varied targets, and the duration of follow-up varied from a few minutes to up to 24 months.

Control Group

The control group care was not clear in some reviews [16,23,24], but others were very specific.

Outcomes

The primary outcomes assessed were clinical outcomes (eg, frequency of hypoglycemic events, symptoms, deaths), surrogate outcomes (eg, glycated hemoglobin [HbA_{1c}], blood pressure, lipid profile, cardiovascular disease risk profiles, lung function tests results, nebulizer use, weight, body mass index [BMI]), behavioral or lifestyle changes (eg, sexual behavior, smoking cessation, increase in physical activity), and processes of care (eg, attendance rates, compliance with medication taking, data management, communication performance, time to diagnosis, time to treatment, changes in professionals' workload).

The secondary outcomes were cost, patient satisfaction, and potential harms and adverse effects.

Quality of Included Studies

Multimedia Appendix 3 shows the results of the quality assessment of the 23 systematic reviews. Ten were rated low quality (AMSTAR score 0-4), seven moderate quality (score 5-8), and six were high quality (score 9-11).

Regarding bias risk assessment, 10 reviews did not explicitly report on study quality assessment. Two specified that the risk of bias was mostly either low or unclear [25]. Free et al [18,19] related that only 4 trials were at low risk of bias in all areas. All studies included in Car et al [13] had high risk of bias.

Baron et al used the McMaster University quality assessment tool and reported that the overall quality was poor [26]. Fanning et al used the Guide to Community Preventative Services data extraction form and reported that 7 studies were rated "fair" and 4 were "good" [16]. Three reviews [27] used the Cochrane Handbook for Systematic Reviews of Interventions [28] reporting varying methodological quality, with some providing insufficient information.

Two reviews reported adequate sequence generation for randomization [20]. De Jongh et al [17] reported adequate sequence generation in 3 of 4 and that none of the included studies were clear if the allocation was concealed. A potential for bias occurred from the apparent lack of blinding of outcome assessors. In Vodopivec-Jamsek et al [27], allocation concealment was considered adequate in 2 studies and unclear in 2 studies. Only one study reported on blinding of personnel collecting and analyzing samples. No mention is made of blinding of outcome assessors or researchers, which could have introduced bias.

Incomplete data analysis methods varied, with analysis and reporting based on intention-to-treat analysis and on only participants who completed the study, which could influence generalizability of the findings. Substantial heterogeneity was detected across analyses; however, a post hoc decision to conduct the main analysis using a random-effects model resulted in no difference in the interpretation of findings. Bacigalupo et al [29] also used Cochrane guidance [28], reporting that 4 of 7 studies presented low risk of bias. The only studies in which Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) methodology [30] was used were those published in the Cochrane Database of Systematic Reviews [13,15,17,20,27].

Main Results

The reviews show a variety of results, as expected, for using different devices and different mHealth interventions on different populations. The most widely used and successful intervention was SMS addressed to chronic disease patients. Positive impact was reported on clinical outcomes, adherence to treatment and care, health behavior changes, disease management, attendance rates, and others. However, some reviews showed conflicting results, with no significant differences between intervention and control groups.

Clinical Outcomes

Asthma

Positive impact has been demonstrated with moderate-quality evidence that text messaging interventions showed greater improvements in the pooled symptom score (mean difference 0.36, 95% CI -0.56 to -0.17) compared with the control group [17]. Very low-quality evidence (GRADE Very Low) showed the following: increased office visits in the SMS group, whereas increased hospital admissions for the control group [17]; a reduction in hospitalizations and better symptom control using home spirometry transmission to physicians via SMS and telephone counseling [25]; increased unscheduled visits to the emergency department and hospitalizations using mobile phone-based interactive self-care software plus management feedback on pulmonary function [25]; and improvements in cough and nighttime symptoms and decreased daily doses of medication using peak flow and symptoms monitoring via SMS [31].

Cardiac Rehabilitation

Exercise capacity in cardiac rehabilitation improved a 6-minute walk test from 524-637 meters (P=.009) in monitored exercise training assisted by a mobile phone app. In an 8-week,

nonrandomized clinical trial, there was 17.6% (SD 16.1) improvement in mobile (n=30) versus 11.5% (SD 35.9) in control group (n=32) (*P*>.05) [32].

Congestive Heart Failure

Mobile technology counseling led to fewer symptom complaints in congestive heart failure subjects [25]. There was relative risk reduction (20%) of death or hospitalization and better quality of life with nurse telephone intervention and cardiologist support [33].

Chronic Lung Diseases

An SMS program improved cough symptoms and sleep quality [34]. Mobile phones recorded respiratory symptoms during exercise training that increased walking distance [34].

Chemotherapy Symptoms

No statistically significant effects were demonstrated on chemotherapy-related toxicity symptoms when patients used a mobile phone app to report symptoms and receive self-care advice [18].

Diabetes

Educational group sessions for diabetic women via SMS showed positive effects on sleep, positive actions, and coping [35].

Surrogate Outcomes

Asthma

There was moderate-quality evidence that text messaging intervention led to greater improvements in peak flow variability (mean difference -11.12, 95% CI -19.56 to -2.68) compared with the control group. No significant differences existed between groups in impact on forced vital capacity or forced expiratory flow in 1 second (GRADE Moderate) [17].

Forced expiratory flow in 1 second improved after 4 months of home spirometry transmission to physicians via SMS and telephone counseling [25]. Using mobile phone-based, interactive self-care software plus management feedback on pulmonary function showed evidence of improvement in pulmonary function and health-related quality of life and decreased unscheduled visits to emergency departments and hospitalizations, plus an increase in the proportion of participants who received leukotriene inhibitors [25].

Cardiac Rehabilitation

Improvement in at least 1 risk factor (relative risk [RR] 1.4, 95% CI 1.1-1.7) in a randomized controlled trial (RCT) that assessed the impact of lifestyle counseling, mobile intervention, devices for home monitoring plus SMS messaging of recommendations, compared to lifestyle counseling alone. The mHealth group was more likely to achieve goals for blood pressure (BP) (62.1% vs 42.9%), HbA_{1c} (86.4% vs 54.2%), and BMI (0.37 kg/m^2decrease vs 0.38 increase). No significant differences in smoking cessation, cholesterol, or medication adherence [32].

Chronic Lung Diseases

Three RCTs showed nonsignificant results in lung function parameters [34].

Diabetes

Educational program via SMS for self-management improved HbA_{1c}, low-density lipoprotein cholesterol, and microalbuminuria [33]. Cloud-based interactive voice response calls and automated email for clinicians reduced HbA_{1c} [33]. SMS promoting medication adherence reduced fasting plasma glucose and 2-hour, postprandial glucose [33]. SMS with tailored instructions on diabetes mellitus care to adolescents and elderly patients improved HbA_{1c} [34]. Diabetes self-management intervention reduced HbA_{1c} [18]. Diabetes education and advice via mobile phone and SMS significantly reduced HbA_{1c} [31]. A mobile phone-based, home glucose monitoring program study decreased HbA_{1c} from 13.2% to 10.5% after 3-6 months [35]. Text messaging improved HbA_{1c}, with positive results in 6 of 8 studies [36]. Daily glucose readings were transmitted via mobile phone to a physician who made adjustments plus clinic appointment [25].

The results were mixed on the impact of mobile telemonitoring supporting diabetes management and feedback on HbA_{1c} but were more consistently positive for studies in type 2 diabetes. Ten of the 13 studies in type 2 diabetes and 4 of 7 studies on type 1 diabetes found mHealth led to HbA_{1c} improvement [26]. Studies without health care professional feedback led to HbA_{1c} improvement, suggesting professional feedback might not be necessary [26]

However, one study showed reduced benefit: educational group sessions for diabetic women via SMS showed higher diastolic blood pressure (+7 mmHg) and less spiritual hope at 6 months, and frequent texters had higher BMI and more sedentary time [35]

HIV Management

Text messages to maintain contact, monitor, and respond to medication issues in patients on antiretrovirals statistically significantly reduced human immunodeficiency virus (HIV) viral load by improving adherence [18].

Hypertension

Improvement in BP has been demonstrated with SMS or voice mail and immediate physician feedback [33]; SMS enabling interactive monitoring, where the provider sets reminders for patients, collects data, and schedules visits [34]; and electronic salt sensor and mobile sensor [34].

However, other reviews showed no benefit. Two trials showed no statistically significant reduction in BP [18]. Groups that did or did not receive alerts and reminders had nearly equal percentages of patients with controlled BP at follow-up [31].

Risk Factors for Coronary Artery Disease

Significant improvements were shown with automatic sphygmomanometer, blood glucose and lipid meter, and mobile phone [34].

Weight Loss

Moderate-quality evidence of short-term weight loss in overweight and obese adults with BMI of 25-39.9 using mHealth structured program was shown [29]. Mobile phone personalized

advice and motivation observed a significant improvement in percent of body fat lost; however, BMI and systolic and diastolic BP were unchanged [31].

Behavioral or Lifestyle Changes

Physical Activity

Seven of 14 trials reported statistically significant benefits on self-reported physical activity outcomes, but no statistically significant change was demonstrated on trials using SMS to reduce calorie intake and increase physical activity or for trials targeting physical activity only, diet only, or diet and physical activity [18].

Smoking Cessation

Positive results have been demonstrated with moderate-quality evidence that mobile phone-based cessation interventions increased abstinence rates at 26 weeks (RR 1.67 95% CI 1.46-1.90, 12 RCTs in high income countries, 11,885 participants, GRADE Moderate). Six studies verified quitting biochemically at 6 months (RR 1.83 95% CI 1.54-2.19) [20]. SMS-based smoking cessation interventions doubled biochemically verified smoking cessation at 6 months [18].

Sexual Behavior

One study showed statistically significant benefits on behavior change [18].

Processes of Care

Antenatal Support

Pregnant women connecting to their health care provider through bidirectional mobile phone messaging were more likely to have skilled birth attendants [37].

Attendance Rates

A consistent improvement on attendance rates has been demonstrated. Text message reminders improved the rate of attendance at health care appointments compared with no reminders (RR 1.10; 95% CI 1.03-1.17) (7 studies, 5841 participants, GRADE Moderate) [15]. They had a similar impact to phone call reminders (RR 0.99, 95% CI 0.95-1.02) (3 studies, 2509 participants, GRADE Moderate). Text messages plus postal reminders improved attendance rates at appointments compared to postal reminders (RR 1.10, 95% CI 1.02-1.19) (1 study, 291 participants, GRADE Low) [13].

In a limited study, Shetty et al [38] compared the effect of one SMS reminder sent to type 2 diabetes patients every third day for 1 year. Although there was no significant reduction in the mean HbA_{1c} values in either group, the percentage of patients with HbA_{1c}<8% decreased significantly in the SMS group. Free et al [18] included one study assessing SMS reminders on attendance for vaccination at different time points (as different studies). The relative risk was 1.19 (95% CI 1.15-1.23), but there was significant heterogeneity (I^2=99.7%, P<.001) [33]. In 8 studies Free et al [19] observed that the pooled effect on appointment attendance using text message reminders versus no reminder was RR 1.06 (95% CI 1.05-1.07) [19]. There was no effect on the number of cancelled appointments (RR 1.08, 95% CI 0.89-1.30), and no difference in attendance using SMS reminders versus other reminders (RR 0.98, 95% CI 0.94-1.02) [19].

Hall et al [22] assessed 6 studies in low- and middle-income countries. The studies showed mHealth interventions to be beneficial, except a pilot study in rural Swaziland [39]. Mahmud et al [40] trained community health workers to use mobile phones for reporting on patient adherence, send appointments reminders, and answer physician queries. This evaluation was based on a retrospective observational study, with the possibility of recall bias.

Data Collection and Health Care Team Communication

Studies that included data collection as a primary mHealth function demonstrated that mobile phones effectively collect and report data, transfer patient-relevant information, and reduce the need for face-to-face communication [37]. There is a reduction in communication delays and improvement on data collection and reporting [23]. One trial reported a statistically significant improvement in nurse/surgeon communication using mobile phones [19].

Adherence to Treatment

mHealth strategies are beneficial to increase adherence to treatment in diabetes patients: SMS to increase adherence to prescriptions in type 2 diabetes [25], electronic blister packs with SMS communication [34], and insulin adherence among type 1 diabetes patients who received tailored text messages with goal-specific prompts [31].

Compliance with medication taking among memory-impaired, HIV-positive patients significantly increased compared to those without impairment. Hepatitis A and B dose vaccination schedules also increased among international travelers with reminders sent to mobile phones [31].

Hall et al reported improved adherence to tuberculosis treatment by using text messages and adherence to HIV therapy, with evidence of reduction of viral load [22], although the authors did not perform quality assessment. The authors commented that a risk-benefit analysis that assessed mHealth reminders to improve tuberculosis medication adherence showed increased mortality and disability-adjusted life years compared to directly observed treatments, and that an RCT in China observed no benefit with voice calls. They also reported limited evidence of contraceptive knowledge improvement with the use of an SMS education scheme; risk reduction of contracting dengue fever, but no significant improvement over alternative schemes; antenatal care improvement, with increases in using skilled birth attendants and women with 4 antenatal visits; and improved vaccination rates in rural Kenya [22].

Hamine et al observed in 2 studies that text messaging tailored to counteract negative beliefs about asthma and education was associated with improved adherence to medication [34].

Diagnosis

Hall et al mentioned studies showing improvement of diagnostic rates of dermatological conditions with mobile teledermatology [17]. However, two trials using mobile phones to transmit photos to offsite clinicians reported significant reductions in correct diagnoses compared to an onsite specialist [18]. One trial

reported reduction in quality of electrocardiography (ECG) transmitted via mobile phone to an ECG transmitted by fax, but with no effects on ECG interpretation [18]. Krishna et al [31] showed that fewer days to diagnosis and treatment were reported among those who were notified of test results via text messages.

Cost

The following studies assessed costs, with evidence of reduction when compared to control groups. SMS reminders were more cost-effective than telephone and were equally efficacious [25]. SMS was found to be 35% and 45% less expensive, respectively, per attendance through reductions in research assistants' work hours and in telecommunications costs. SMS reminders were less expensive than mobile phone reminders [33,40]. The relative cost of the text message per attendance was 55% and 65% of the cost of phone call reminders [13]. There was a reduction in patient burden to transportation time and costs in African countries [23]. The cost of text-based support per 1000 enrolled smokers was GB 278 per quitter, but when future health service costs were included, text-based support was considered a cost-saver [20].

Patient Satisfaction

The individual studies did not assess this outcome.

Potential Harms and Adverse Effects

Only 3 reviews assessed this outcome [15,17,27]. Vodopivec-Jamsek et al [27] reported one study where mobile phone messaging was used to support smoking cessation and that messaging did not have any impact on rates of pain in the thumb or finger joints (RR 1.08, 95% CI 0.74-1.59), or car crash rates (RR 0.88, 95% CI 0.58-1.35) at 26 weeks follow-up.

Discussion

Principal Findings

The current evidence shows benefits of mHealth in chronic disease management, improving symptoms and peak flow variability in asthma patients and chronic pulmonary diseases symptoms; heart failure symptoms, reducing deaths and hospitalization and improving quality of life; glycemic control in diabetes patients; and improving BP in hypertensive patients. SMS reminders improved attendance rates at reduced costs and improved adherence to tuberculosis and HIV therapy in some scenarios, with evidence of decrease of viral load.

Mobile devices may improve patient-provider communication, facilitating assistance in disease management. It may increase the likelihood of delivering health interventions to hard-to-reach populations. Whittaker et al [20] listed advantages of using mobile interventions: convenience, ease, cost-effectiveness, scalability, personalization, and "the ability to send time-sensitive messages with an 'always on' device." Hamine et al [34] observed that mHealth tools can impact patients who are less inclined to engage with traditional health services. Aranda-Jan et al [23] reported that governments may benefit from increased support of patient management and increased direct communication in rural areas. Health workers may receive support through professional networks and can prioritize efforts and increase their role in active case detection using disease surveillance systems. Baron et al [26] assessed studies involving data transferring for diabetes management and suggested that recording and tracking of data might increase patients' motivation to self-manage.

The most popular mHealth intervention was behavior change interventions using text messaging. The low cost and low broadband requirements facilitate the spread of applications, even in low-income countries.

Different uses of motivation have also been described as a tool to be used in mHealth interventions in some of the systematic reviews analyzed. These are mainly focused on patient motivation in different contexts on chronic diseases [36], communicable diseases [23], physical activity [16,29,31], and empowerment in the use of services [37].

Two reviews [14,32] reported lessons learned: patient needs must be met, training and support provided, users engaged in the development and implementation of the tools, and consideration of patient age and education level. Usage might improve with user-centered design, engagement strategies, and feedback to the users.

This study provides a thorough review of available evidence on effectiveness of mHealth interventions in different health conditions and in the processes of health care service delivery, so it useful to guide clinical and health policy decisions. However, there are some limitations of the studies that need to be addressed.

Studies assessing mHealth interventions usually do not include the assessment of risks, consumer satisfaction, and acceptability of the intervention [17]. None reported studies assessing security and confidentiality. Chen et al [41] noted that mobile phone numbers frequently change in China, reducing certainty the message was delivered to the correct recipient. This was not assessed elsewhere. Particularly in low-income countries where mobile phones are frequently shared, these are important confidentiality issues that must be taken into account when designing interventions [15]. De Jongh et al [23] warned of inaccurate data input, misinterpretation of the information, and difficulties in reading due to vision or literacy problems and remarked that text messaging cannot capture verbal and nonverbal cues. Norwell suggested that doctors agree on vocabulary to minimize the risk of patients' misunderstanding the message [38]. Risks associated with mobile phone messaging in general may apply, such as car accidents [23].

Other drawbacks related to mHealth initiatives were reported by Hamine et al [34] Some patients' concerns included dependence on professional supervision, unnecessary medicalization, fear of technology failure, and difficulty in understanding and using the technology. Provider concerns related to data review and response times, increased clinical workload and workflow, record maintenance, and concerns about supervision and technology dependence. Aranda-Jan et al [23] reported difficulties in monitoring text message content, data underreporting, and the possibility of biased responses from participants.

Two reviews cite availability and poor connectivity as barriers [34]. Most identified the main limitation as the small number

of RCT studies, patients enrolled, and the low-to-moderate quality of evidence. Researchers should validate their pilot study findings through follow-up studies with adequate research designs and appropriate controls [17]. Aranda-Jan et al [23] mention that claimed benefits are unclear and long-term results uncertain. Also, only 2 reviews assessed funding [15]. This is important to identifying conflicts of interest. To improve the suboptimal reporting and standardize Web-based and mobile health interventions, the CONSORT-EHEALTH was developed, a checklist that is an extension of the CONSORT statement [42].

Costs have not been routinely assessed. Such costs may be dependent on the nature of the intervention and the size and characteristics of the target group [17]. More attention to cost implications seems warranted [27]. Additionally, future studies should compare effects in different contexts [27].

Conclusion

Although mHealth is growing in popularity, the evidence for efficacy is still limited. Positive results were reported for chronic disease management, improving chronic pulmonary diseases symptoms and heart failure symptoms, reducing deaths and hospitalization and improving quality of life, and improving glycemic control in diabetes patients and BP in hypertensive patients. SMS reminders improved attendance rates and improved adherence to tuberculosis and HIV therapy in some scenarios. However, in general the methodological quality of the studies included in the systematic reviews is low. For some fields, its impact is not evident or is mixed. Exceptions are the moderate improvement in asthma patients, attendance rates, and smoking cessation rates.

Acknowledgments

The authors thank the team that worked on the systematic review: Renato Minelli Figueira, Gustavo Moreira Alkmim, Maria Clara Noman de Alencar, Nayara Dornela Quintino, Elsy Dumit-Bechara, Francesc Saigí-Rubió, Francisco Becerra-Posada, Edward Talbott Kelley and the undergratuate students Edson Alexandre Silva Carvalho, Fernanda Cotrim Stefanelli, Gabriel Almeida Silqueira Rocha, Grace Kelly Matos e Silva, Marcelo Henriques de Camargos, Thales Matheus Mendonça Santos, Thiago Adriano de Deus Queiroz Santos.

DNO and MD are staff members of the Pan American Health Organization / World Health Organization (PAHO/WHO) and are solely responsible for the views expressed in the paper, which does not necessarily represent the views, decisions, or policies of the PAHO/WHO.

Authors' Contributions

DNO, MBMA, and ALR conceptualized the study and designed the review. MSM and JAQO designed the search strategies with input from DNO, MBMA, and ALR. MBMA and ALR randomly checked the accuracy of extracted data. MSM and JAQO wrote the first draft of the review paper. MBMA, MD, ALR, and DNO contributed to the final draft of this manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search methods and strategy.

[PDF File (Adobe PDF File), 23KB - mhealth_v6i1e23_app1.pdf]

Multimedia Appendix 2

Descriptive summary of the 23 systematic reviews included in mHealth review.

[PDF File (Adobe PDF File), 172KB - mhealth_v6i1e23_app2.pdf]

Multimedia Appendix 3

Quality assessment ratings of systematic reviews included in mHealth review.

[PDF File (Adobe PDF File), 76KB - mhealth_v6i1e23_app3.pdf]

References

1. Carter A, Liddle J, Hall W, Chenery H. Mobile Phones in Research and Treatment: Ethical Guidelines and Future Directions. JMIR Mhealth Uhealth 2015 Oct 16;3(4):e95 [FREE Full text] [doi: 10.2196/mhealth.4538] [Medline: 26474545]

2. Donner J. Research Approaches to Mobile Use in the Developing World: A Review of the Literature. The Information Society 2008 May 06;24(3):140-159. [doi: 10.1080/01972240802019970]
3. World Health Organization. New horizons for health through mobile technologies. Geneva: World Health Organization; 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2018-01-09] [WebCite Cache ID 6wLiS9YZp]
4. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
5. Nglazi MD, Bekker L, Wood R, Hussey GD, Wiysonge CS. Mobile phone text messaging for promoting adherence to anti-tuberculosis treatment: a systematic review. BMC Infect Dis 2013 Dec 02;13:566 [FREE Full text] [doi: 10.1186/1471-2334-13-566] [Medline: 24295439]
6. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: An updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
7. Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008 May 20;148(10):776-782. [Medline: 18490690]
8. Silva C, Zamboni A, Hernandes E, Di Thomazzo A, Belgamo A, Fabbri S. LaPES. StArt - State of the Art through Systematic Review URL: http://lapes.dc.ufscar.br/tools/start_tool [accessed 2018-01-10] [WebCite Cache ID 6wNIn2Oox]
9. Smith V, Devane D, Begley CM, Clarke M. Methodology in conducting a systematic review of systematic reviews of healthcare interventions. BMC Med Res Methodol 2011 Feb 03;11(1):15 [FREE Full text] [doi: 10.1186/1471-2288-11-15] [Medline: 21291558]
10. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007 Feb 15;7:10 [FREE Full text] [doi: 10.1186/1471-2288-7-10] [Medline: 17302989]
11. Finitsis DJ, Pellowski JA, Johnson BT. Text message intervention designs to promote adherence to antiretroviral therapy (ART): a meta-analysis of randomized controlled trials. PLoS One 2014;9(2):e88166 [FREE Full text] [doi: 10.1371/journal.pone.0088166] [Medline: 24505411]
12. Horvath T, Azman H, Kennedy GE, Rutherford GW. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection. Cochrane Database Syst Rev 2012 Mar 14(3):CD009756. [doi: 10.1002/14651858.CD009756] [Medline: 22419345]
13. Car J, Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2012 Jul 11(7):CD007458. [doi: 10.1002/14651858.CD007458.pub2] [Medline: 22786507]
14. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]
15. Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R, Car J. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2013 Dec 05(12):CD007458. [doi: 10.1002/14651858.CD007458.pub3] [Medline: 24310741]
16. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
17. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
18. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
19. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
20. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
21. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]
22. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7:25606 [FREE Full text] [Medline: 25361730]

23. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
24. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
25. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
26. Baron J, McBain H, Newman S. The impact of mobile monitoring technologies on glycosylated hemoglobin in diabetes: a systematic review. J Diabetes Sci Technol 2012 Sep 01;6(5):1185-1196 [FREE Full text] [doi: 10.1177/193229681200600524] [Medline: 23063046]
27. Vodopivec-Jamsek V, de Jongh T, Gurol-Urganci I, Atun R, Car J. Mobile phone messaging for preventive health care. Cochrane Database Syst Rev 2012 Dec 12;12:CD007457. [doi: 10.1002/14651858.CD007457.pub2] [Medline: 23235643]
28. Higgins J, Green S. The Cochrane Collaboration. 2011. Cochrane Handbook for Systematic Reviews of Interventions URL: http://training.cochrane.org/handbook [accessed 2018-01-09] [WebCite Cache ID 6wLih6wy5]
29. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley WH. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291 [FREE Full text] [doi: 10.1111/obr.12006] [Medline: 23167478]
30. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
31. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
32. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
33. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
34. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
35. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014 Jun 13;10:49 [FREE Full text] [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]
36. Valles Ortiz P, Miranda Félix P, García Sosa E. Mensajes de texto para el control glucémico en adultos con diabetes tipo 2: revisión sistemática. eGlobal 2015 Jan 01;14(1):435-444 [FREE Full text] [doi: 10.6018/eglobal.14.1.188461]
37. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
38. Shetty AS, Chamukuttan S, Nanditha A, Raj RKC, Ramachandran A. Reinforcement of adherence to prescription recommendations in Asian Indian diabetes patients using short message service (SMS)--a pilot study. J Assoc Physicians India 2011 Nov;59:711-714. [Medline: 22616337]
39. Kliner M, Knight A, Mamvura C, Wright J, Walley J. Using no-cost mobile phone reminders to improve attendance for HIV test results: a pilot study in rural Swaziland. Infect Dis Poverty 2013 Jun 14;2(1):12 [FREE Full text] [doi: 10.1186/2049-9957-2-12] [Medline: 23849508]
40. Mahmud N, Rodriguez J, Nesbit J. A text message-based intervention to bridge the healthcare communication gap in the rural developing world. Technol Health Care 2010;18(2):137-144. [doi: 10.3233/THC-2010-0576] [Medline: 20495253]
41. Chen Z, Fang L, Chen L, Dai H. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: a randomized controlled trial. J Zhejiang Univ Sci B 2008 Jan;9(1):34-38 [FREE Full text] [doi: 10.1631/jzus.B071464] [Medline: 18196610]
42. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

AMSTAR: Measurement Tool to Assess Systematic Reviews
BP: blood pressure

BMI: body mass index
CONSORT: Consolidated Standards of Reporting Trials
ECG: electrocardiography
GRADE: Grading of Recommendations, Assessment, Development and Evaluation
HbA_{1c}: glycated hemoglobin
IBECS: Indice Bibliográfico Español de Ciencias de la Salud
LILACS: Literatura Latino-Americana e do Caribe em Ciências da Saúde
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: relative risk
SD: standard deviation
SMS: short message service

Edited by G Eysenbach; submitted 30.08.17; peer-reviewed by W Curioso, A Benis; comments to author 26.09.17; revised version received 21.11.17; accepted 08.12.17; published 17.01.18.

Please cite as:
Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D
The Impact of mHealth Interventions: Systematic Review of Systematic Reviews
JMIR Mhealth Uhealth 2018;6(1):e23
URL: http://mhealth.jmir.org/2018/1/e23/
doi:10.2196/mhealth.8873
PMID:29343463

Original Paper

Smartphone App–Based Assessment of Gait During Normal and Dual-Task Walking: Demonstration of Validity and Reliability

Brad Manor[1], PhD; Wanting Yu[1], MEng; Hao Zhu[1], MSc; Rachel Harrison[1], BS; On-Yee Lo[1], PhD; Lewis Lipsitz[1], MD; Thomas Travison[1], PhD; Alvaro Pascual-Leone[2], MD, PhD; Junhong Zhou[1], PhD

[1]Hebrew SeniorLife Institute for Aging Research, Harvard Medical School, Roslindale, MA, United States

[2]Berenson-Allen Center for Noninvasive Brain Stimulation, Division of Interventional Cognitive Neurology, Beth Israel Deaconess Medical Center, Boston, MA, United States

Corresponding Author:
Brad Manor, PhD
Hebrew SeniorLife Institute for Aging Research
Harvard Medical School
1200 Centre St
Roslindale, MA,
United States
Phone: 1 617 971 5332
Email: bradmanor@hsl.harvard.edu

Abstract

Background: Walking is a complex cognitive motor task that is commonly completed while performing another task such as talking or making decisions. Gait assessments performed under normal and "dual-task" walking conditions thus provide important insights into health. Such assessments, however, are limited primarily to laboratory-based settings.

Objective: The objective of our study was to create and test a smartphone-based assessment of normal and dual-task walking for use in nonlaboratory settings.

Methods: We created an iPhone app that used the phone's motion sensors to record movements during walking under normal conditions and while performing a serial-subtraction dual task, with the phone placed in the user's pants pocket. The app provided the user with multimedia instructions before and during the assessment. Acquired data were automatically uploaded to a cloud-based server for offline analyses. A total of 14 healthy adults completed 2 laboratory visits separated by 1 week. On each visit, they used the app to complete three 45-second trials each of normal and dual-task walking. Kinematic data were collected with the app and a gold-standard–instrumented GAITRite mat. Participants also used the app to complete normal and dual-task walking trials within their homes on 3 separate days. Within laboratory-based trials, GAITRite-derived heel strikes and toe-offs of the phone-side leg aligned with smartphone acceleration extrema, following filtering and rotation to the earth coordinate system. We derived stride times—a clinically meaningful metric of locomotor control—from GAITRite and app data, for all strides occurring over the GAITRite mat. We calculated stride times and the dual-task cost to the average stride time (ie, percentage change from normal to dual-task conditions) from both measurement devices. We calculated similar metrics from home-based app data. For these trials, periods of potential turning were identified via custom-developed algorithms and omitted from stride-time analyses.

Results: Across all detected strides in the laboratory, stride times derived from the app and GAITRite mat were highly correlated (P<.001, r^2=.98). These correlations were independent of walking condition and pocket tightness. App- and GAITRite-derived stride-time dual-task costs were also highly correlated (P<.001, r^2=.95). The error of app-derived stride times (mean 16.9, SD 9.0 ms) was unaffected by the magnitude of stride time, walking condition, or pocket tightness. For both normal and dual-task trials, average stride times derived from app walking trials demonstrated excellent test-retest reliability within and between both laboratory and home-based assessments (intraclass correlation coefficient range .82-.94).

Conclusions: The iPhone app we created enabled valid and reliable assessment of stride timing—with the smartphone in the pocket—during both normal and dual-task walking and within both laboratory and nonlaboratory environments. Additional work is warranted to expand the functionality of this tool to older adults and other patient populations.

(JMIR Mhealth Uhealth 2018;6(1):e36) doi:10.2196/mhealth.8815

KEYWORDS

smartphone; gait assessment; pocket; dual task; validity; reliability; mobile applications

Introduction

Walking is central to many activities of daily living and is most typically completed while simultaneously performing unrelated cognitive tasks, for example, talking, reading signs, or making decisions. Even in healthy adults, such dual tasking reduces gait speed, prolongs stride time, and increases stride-to-stride movement variability [1]. These performance decrements, or "costs," indicate that walking is regulated by a complex control system dependent on numerous cognitive functions and underlying brain networks. Therefore, the assessment of gait under normal and dual-task conditions provides valuable insights into not only one's physical health [2,3], but also one's brain health, and even the likelihood of developing dementia several years into the future [4,5].

Gait assessments are typically completed within clinical or laboratory settings. They are thus inaccessible to those living in remote settings and do not lend themselves well to high-frequency monitoring. Moreover, clinical assessments entail qualitative evaluation, are predisposed to subjective bias, and are often insensitive to subtle gait disturbances [6-9]. Laboratory assessments overcome these limitations by quantifying temporospatial characteristics of gait, yet they require expensive equipment, dedicated laboratory space, and trained personnel. There is thus an urgent need to develop mobile tools that enable low-cost quantitative assessments of gait.

Smartphones contain a 3-dimensional accelerometer, a 3-dimensional gyroscope, and a digital compass that are similar in sensitivity to research-grade biomechanical instrumentation. The smartphone, when secured to an individual's lower back or sternum as they walk, can detect gait events such as heel strikes [10], as well as kinematic differences between those with and those without movement disorders, such as Parkinson disease [11,12]. Still, studies to date have been limited to laboratory environments and have required trained personnel to administer assessments, provide instructions, and secure the phone to the participant's trunk.

In collaboration with Sage Bionetworks (Seattle, WA, USA) and supported by the Football Players Health Study at Harvard University, the objective of this study was to create an iPhone-based app enabling the administration of a standardized gait assessment, under both normal and dual-task conditions, within nonlaboratory settings. The app was designed to provide multimedia instructions to the user, acquire data with the phone placed in the user's pocket, and derive stride times from bouts of walking. We chose stride time because it can be directly derived from gait events (eg, heel strikes), is closely linked to gait speed [13], and has been associated with aging [14], movement disorders [15], cognitive impairment [16], and the development of falls [17]. As turning significantly disrupts stride timing [18], we also developed a method of automatically detecting turns. We determined the validity and reliability of the app by (1) comparing the accuracy of stride times derived from the app versus those derived from gold-standard laboratory instrumentation, and (2) determining the test-retest reliability of app-derived stride times within both laboratory- and home-based settings.

Methods

Smartphone App

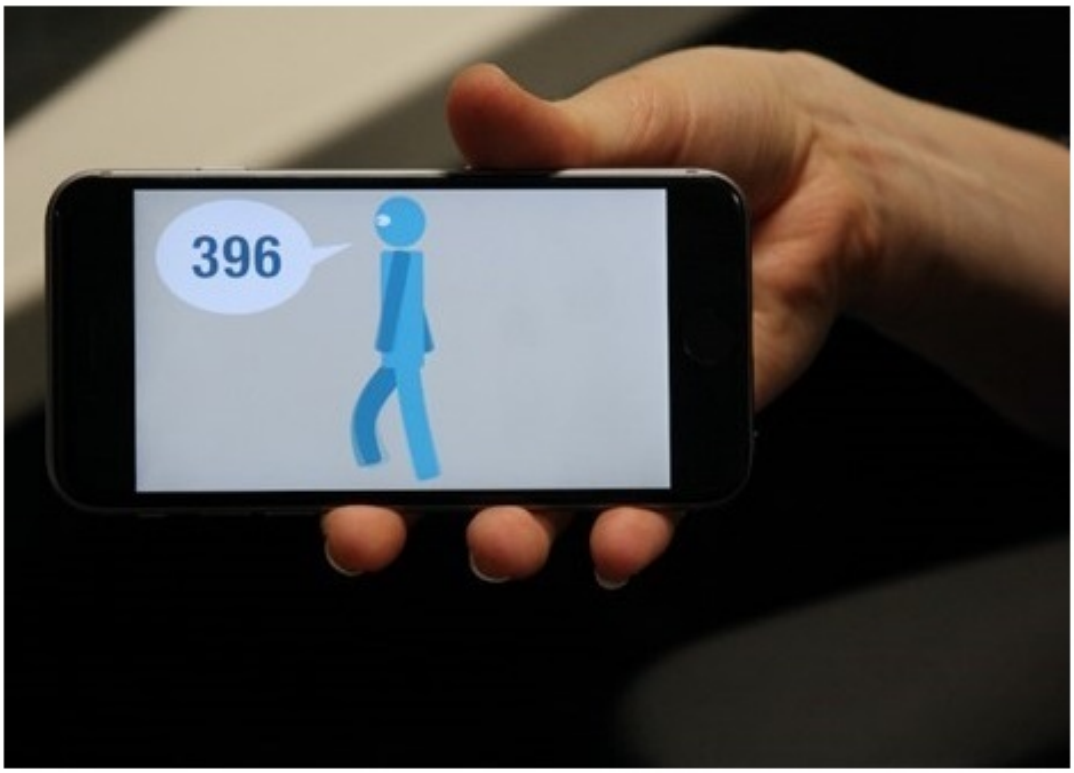
The app was designed to recreate a common laboratory-based dual-task gait assessment, namely, evaluation of walking under normal conditions and again while verbalizing serial subtractions of 3 from a random number between 200 and 999 [19,20]. The app provided multimedia instructions to the participant to help ensure reliability of results. Participants first watched an animation developed by Wondros Inc (Los Angeles, CA, USA) that provided a general overview of the assessment (Figure 1). The user was then presented with several on-screen text instructions. The last page instructed the participant to press Start and place the phone in their preferred front pocket. The iPhone speaker was then used to provide audible instructions to the participant for the remainder of the assessment. These instructions provided the procedural details of each walking trial, cues for the start and end of each trial, and, for dual-task trials, a randomly generated starting number for the serial-subtraction task.

We designed walking assessments to include one 45-second trial of normal walking and one 45-second trial of dual-task walking. Trial start and end cues triggered acquisition of accelerometer, gyroscope, and magnetometer data, which were stored on the phone's internal storage capacity. Following each assessment, the participant was prompted to answer a multiple-choice question (see following section) presented in text format on the smartphone screen. Kinematic and questionnaire data were then automatically uploaded via Wi–Fi to a remote, cloud-based data server for offline analyses.

Participants

We recruited men and women aged 18 to 35 years via local advertisement. Exclusion criteria were an inability to walk unassisted; self-report of major disease, such as stroke, Parkinson disease, diabetes mellitus, or cardiovascular disease; history or presence of ulceration, amputation, or other painful symptoms in the lower extremities; drug or alcohol abuse; and hospitalization within the past 6 months. Interested and eligible individuals provided written informed consent as approved by the Hebrew SeniorLife Institutional Review Board (Hebrew SeniorLife Institute for Aging Research, Roslindale, MA, USA; approval number: IRB-2015-40).

Figure 1. Screenshot of the animated instruction for dual-task walking. The app provides text and animated instructions prior to the assessment, followed by voice instructions during the assessment, to enable gait analysis from data acquired while the user walks with their phone placed in the pocket of their pants or shorts. Gait is assessed while individuals walk normally, and again while they walk and simultaneously perform a serial-subtraction cognitive dual task.



Study Procedures

Participants completed 2 laboratory visits separated by 1 week. They completed the same assessments of locomotor control within each visit, during which data were simultaneously collected via the app and a 14-foot instrumented GAITRite mat (CIR Systems, Inc, Franklin, NJ, USA). Participants additionally used the app to complete walking assessments within their homes on 3 separate days, in between their 2 laboratory visits. No instructions were provided regarding time of day for home assessment completion.

Laboratory Assessments

Participants completed 2 laboratory visits separated by at least one week. We instructed them to wear comfortable shoes and pants or shorts with front pockets for each visit. The same procedures were completed on each visit to enable testing of between-visit test-retest reliability. Within each visit, participants completed the app walking assessment 3 separate times, such that they completed 3 pairs of normal walking and dual-task walking trials. Trial order was randomized with each pair.

Within the laboratory, each walking trial was completed around an oval-shaped, 24-m indoor track. We placed the GAITRite mat along one long side of the track. Each trial began with participants standing just behind the beginning of the mat to ensure that the first footfall of each trial was captured by the mat. Participants used the app instructions to initiate and complete each trial.

After all trials, the app prompted participants to use the iPhone touch screen to answer the following multiple-choice question: “How tight is the pocket in which you placed the phone? (tight, medium or loose).” We did this to study the effects of this variable on the ability to collect valid and reliable data over time. The questionnaire was incorporated into the app using the SageBridge online portal (Sage Bionetworks).

Home Assessments

We asked participants to use the app to complete a walking assessment (1 normal walk and 1 dual-task walk) at home on 3 separate days in between their laboratory visits. The app provided the same instructions to the participant as during the laboratory visit. Additionally, the app instructed participants to complete the walk in a quiet room or hallway, and to walk continuously throughout the trial, making turns if and when needed. On completion of both trials, participants were prompted to answer the same multiple-choice question regarding pocket tightness as described in the laboratory assessment.

Data Analysis

Laboratory Assessments

The app sampled kinematic data at a frequency of 100 Hz. Raw, 3-dimensional accelerometer and gyroscope time series were each transformed from the device coordinate system to an earth coordinate system using the quaternion rotation matrix.

Following this rotation, the z-axis formed a line between the center of the earth and the phone, and was thus approximately vertical to the ground (see Figure 2, part A, for example acceleration data). Each rotated z-axis time series was then filtered with a common Butterworth filter. These time series, which contained peaks that alternated between relatively high and low amplitudes, aligned with heel-strike and toe-off events derived from the GAITRite mat (Figure 2, part B). Specifically, each heel strike corresponded to the trough nadir following each relatively high peak, whereas toe-offs corresponded to the trough nadir following each relatively low peak.

We defined stride time as the time elapsed between 2 consecutive heel strikes of the same foot. We calculated it by determining the number of data points between 2 consecutive trough nadirs following relatively high peaks, and then dividing by the sampling frequency of 100 Hz. For all strides that took place on the GAITRite mat, we calculated stride times from both gait mat and app data and used these for analyses.

Automatic Turn Detection

Walking trials completed at home likely included variable amounts of turning. Turning, while itself an important functional measure, alters stride timing [18]. We therefore developed a method to identify relatively sharp, rapid turns to enable stride-time calculation from bouts of walking without such turns. Turning produces a large deviation in the angular velocity about the body's vertical axis. In pilot studies of straight-line walking with the phone placed in the pocket, z-axis gyroscope data contained fluctuations of relatively small amplitude with frequent zero crossings (Figure 3, part A, black line portions). However, during a 180° turn, this angular velocity was significantly greater in 1 direction—depending on the direction of turning relative to the phone's orientation in the pocket—with no zero crossings (see Figure 3, part A, red line portion).

The total angular distance traveled in 1 direction can be calculated by integrating the angular velocity time series between 2 consecutive zero crossings (ie, area under the curve [AUC]). A 180° turn of the phone's gyroscope would thus equal 3.14 radians (ie, π). The AUC related to the 180° turn depicted in Figure 3, part A, was 3.37. We also noted in our unpublished pilot studies that relatively rapid turns were less likely to contain higher-frequency angular velocity fluctuations that crossed zero. For this analysis, we therefore defined a turn as any period between 2 consecutive zero crossings within rotated, filtered z-axis angular velocity time series in which the product of the AUC and the time between zero crossings eclipsed a predefined threshold.

Figure 2. Example of (A) raw and (B) filtered smartphone-recorded accelerations along the earth coordinate system vertical axis during straight walking, relative to identified gait events. Phone-side leg heel-strike and toe-off events derived from a GAITRite mat were overlaid on vertical-axis accelerations acquired by a smartphone placed in the participant's pocket. These heel-strike and toe-off events correspond to trough nadirs following relatively high peaks and low peaks, respectively, within the filtered acceleration time series.

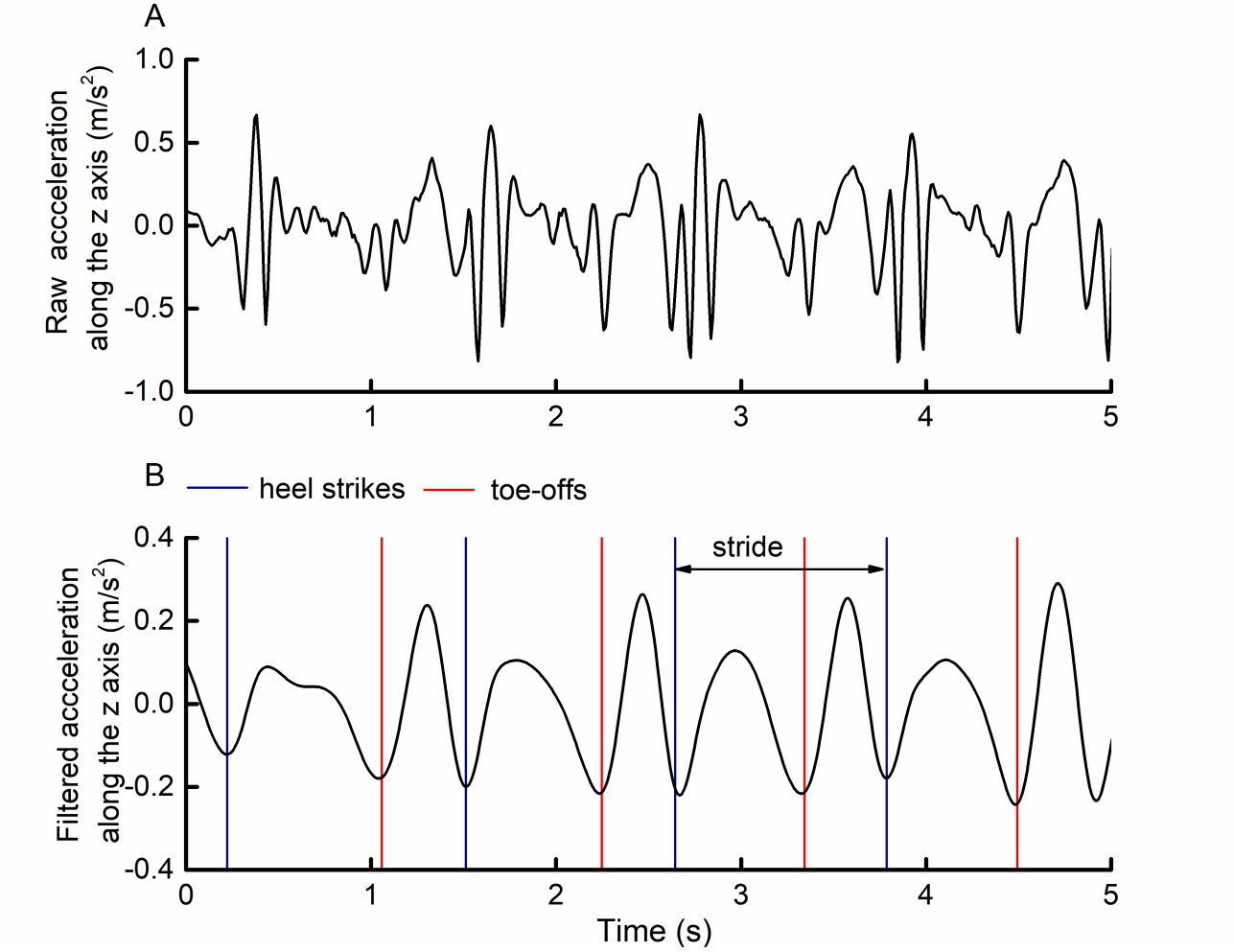


Figure 3. Example of smartphone-recorded (A) angular velocity and (B) acceleration relative to the earth coordinate system vertical axis during straight walking and a 180° turn. In a pilot trial, a participant walked straight across the laboratory before turning around a cone. Angular velocities were relatively small and contained numerous zero crossings during the straight-walking portion of the trial. Turning, on the other hand, was associated with a large nonstationarity in angular velocity between the 2 adjacent zero crossings. Acceleration patterns were noticeably altered during this period. This observation was subsequently used to develop a method to identify potential turns from walking trials collected during in-home assessments.

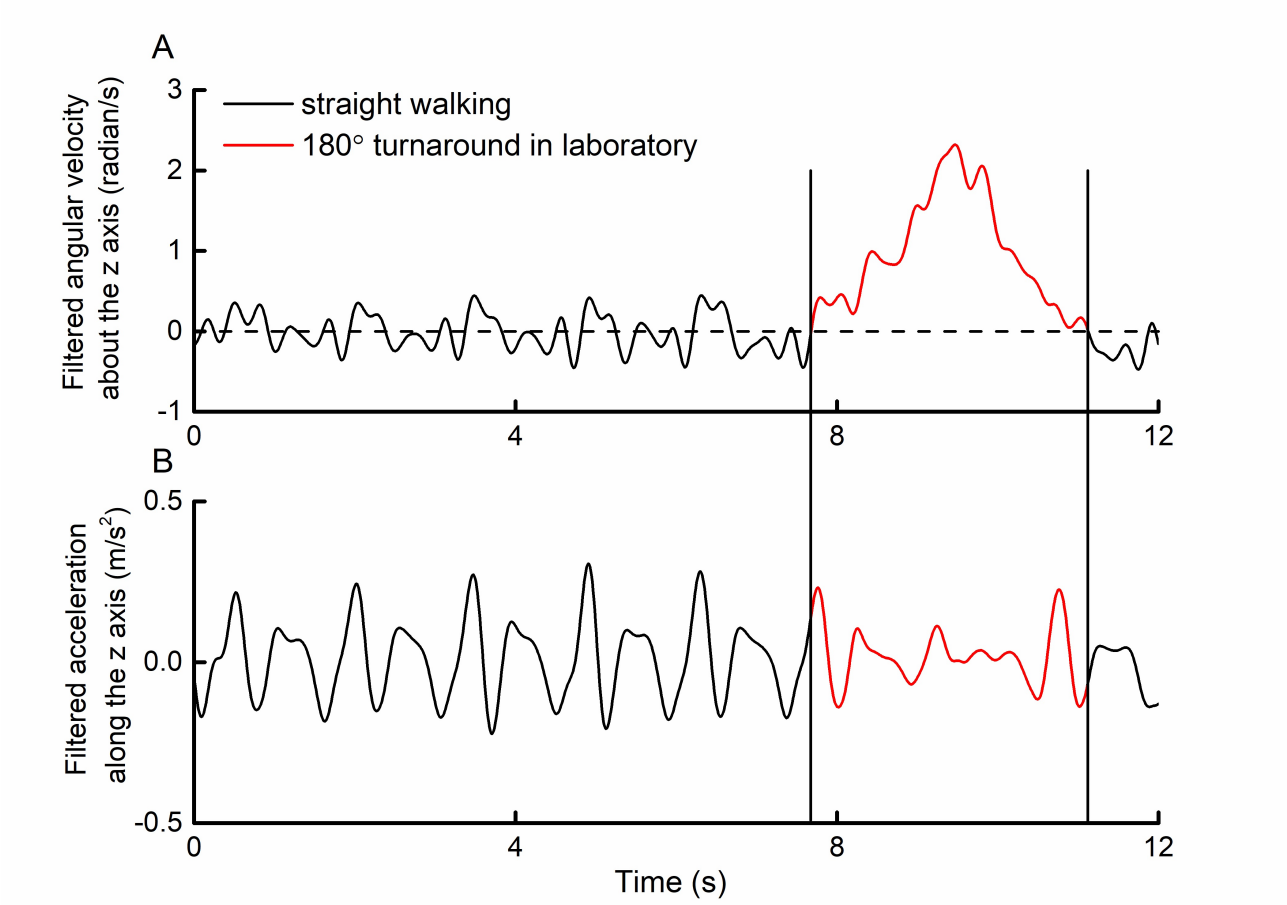


Figure 4 illustrates an example of 2 identified turns within a selected trial of walking within the home. For all home walking trials, we first detected all potential turns as described above. We then calculated stride times from all z-axis acceleration data that occurred outside of detected periods of turning (using the same methods as those described in the "Laboratory Assessments" subsection.

Statistical Analysis

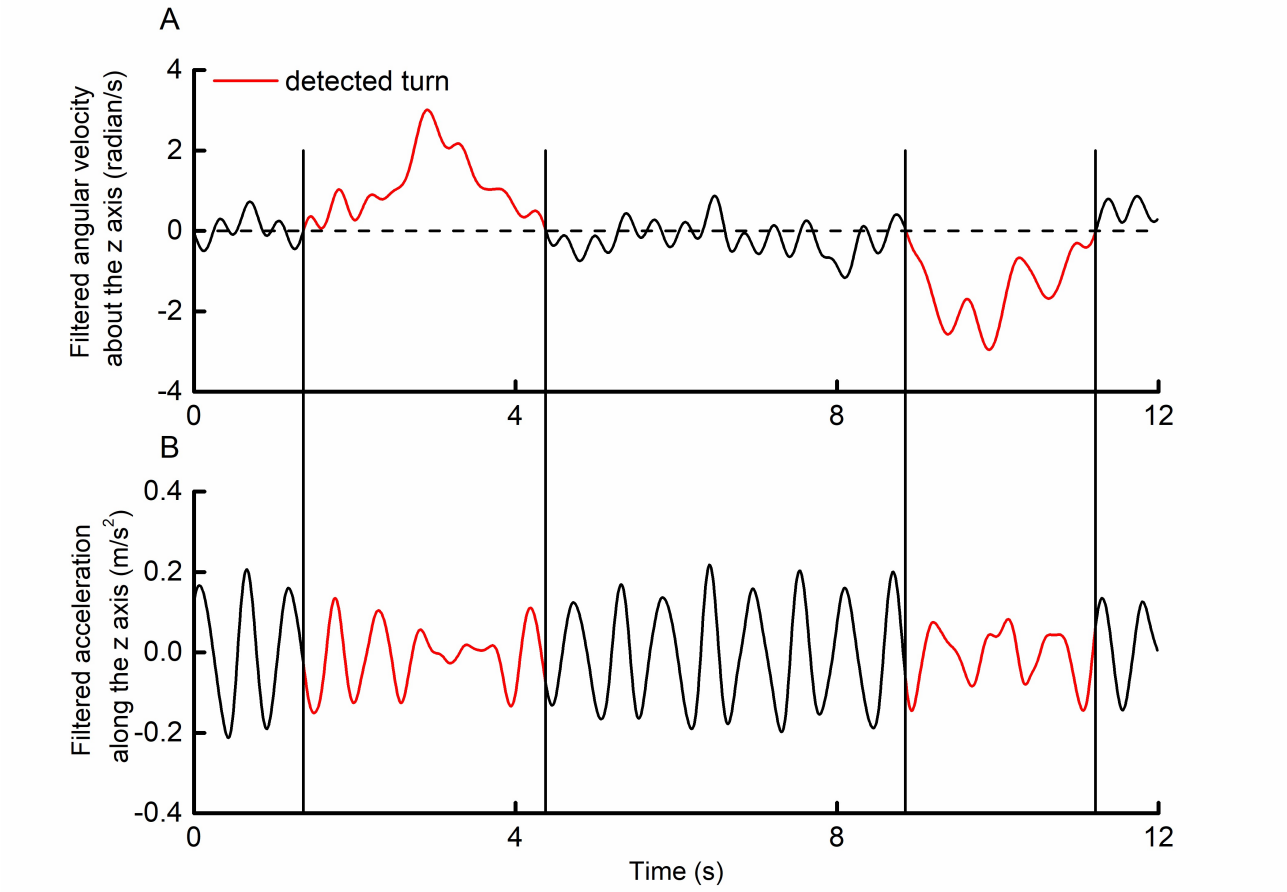
We performed the following analyses with JMP Pro 13 (SAS Institute) and R version 3.3.1 (R Foundation). We set the significance level for all tests to $P<.05$.

Validity

We examined the validity of app-derived stride times by first assessing their agreement with corresponding stride times derived from the GAITRite mat using a Passing-Bablok orthogonal regression model, an appropriate approach for comparing methods while acknowledging measurement error [21]. Models included every stride that occurred during the first pass over the mat, for all trials of both normal and dual-task walking for each participant. We further examined the relationship between individual stride times derived from each device using linear regression and included visit (laboratory visits 1 and 2), task condition (normal or dual-task walking), and pocket tightness (tight, medium, or loose) as model effects to determine whether these factors influenced the observed relationship between derived stride times. In this model, we included participant as a random effect variable, as each contributed multiple data points to the model. We also used similar orthogonal and linear regression models to assess the agreement between dual-task costs to stride time as measured by the app and the GAITRite mat. We calculated cost from each pair of normal and dual-task walking trials as the percentage change (ie, increase) in average stride time.

For each individual stride that occurred on the GAITRite mat, we then calculated the magnitude of error between its stride time as calculated by the app and the GAITRite mat. We produced a Bland-Altman plot of this error to visualize this error as a function of stride time (ie, the average of the individual stride time as calculated by the app and the GAITRite mat). We used 2-way ANOVA to determine whether the magnitude of error, at the individual stride level, was influenced by task condition or pocket tightness.

Figure 4. Example of smartphone-recorded (A) angular velocity and (B) acceleration relative to the earth coordinate system vertical axis during in-home walking with 2 detected periods of turning. Turns were identified from angular velocity time series and defined as any period between 2 consecutive zero crossings in which the product of the area under the curve and the time between zero crossings was >2.00 radian-seconds. Acceleration patterns were noticeably different during these periods. Average stride times from home assessment trials were thus computed from stride times derived only from nonturning periods.



Reliability

We examined the test-retest reliability of the app assessment by computing several intraclass correlation coefficients (ICCs). We calculated ICCs separately for normal and dual-task walking trials, for each of the following 4 conditions: (1) across trials within each laboratory assessment, (2) across trials over the 3 home assessments, (3) between the 2 laboratory visits, and (4) between laboratory and home assessments. For conditions 1 and 2, the unit of interest was the average stride time derived from each trial (ICC 1, 1). For conditions 3 and 4, the unit of interest was the average stride time, averaged across all trials of the same condition (ie, normal or dual task; ICCs 1, 3). We took ICC values greater than .80 to reflect excellent test-retest reliability.

Effects of Participant Characteristics, Walking Condition, and Testing Setting

We used Pearson correlations to examine relationships between average stride times and participant height and body mass. We used 2-way ANOVAs to examine the effects of walking condition (normal walking, dual tasking), setting (laboratory, home), and their interaction on average stride time. Significance level was set to $P<.05$.

Results

We recruited 14 healthy participants aged 22 to 35 years (8 female; mean age 29.6, SD 4.2 years; mean height 168, SD 12 cm; mean body mass 76, SD 14 kg). All 14 participants completed both laboratory visits and all 3 home assessments. Across all 69 recorded assessments, 10 were completed with self-report of loose-fitting pockets, 24 with tight-fitting pockets, and 35 with pockets of medium tightness. For the 41 assessments completed at home, 11 were conducted in the morning, 13 in the afternoon, and 17 in the evening. Across participants, the average day-to-day variation in the range of timing of the 3 home assessments was 4.0 (SD 4.0) hours. The average number of strides detected in each 45-second home-based trial, after removal of turns, was 21.6 (SD 6.4) (range 13-28 strides).

Validity of Smartphone-Derived Stride Time

For each detected stride across all participants and laboratory trials, stride times derived from the app demonstrated excellent validity as compared with the GAITRite mat. Orthogonal regression analysis revealed that stride times derived from the app were highly correlated with those measured by the GAITRite mat ($P<.001$, $r^2=.98$; Figure 5).

Figure 5. Correlation and agreement between stride times derived from a smartphone placed in the pocket and a GAITRite mat. (A) The timing of each individual stride that occurred over the GAITRite mat during all trials of normal and dual-task walking over 2 visits are presented separately for each participant. Stride times were noticeably longer during dual-task walking than during normal walking, for multiple participants. The gray background plot in each subplot is the same and represents the entire sample of stride times. Stride times derived from the app and the GAITRite mat were strongly correlated with one another (r^2=.98, P<.001). The orthogonal best fit line of this entire sample had a slope of approximately 1 and an intercept of approximately 0. (B) Bland-Altman scatterplot depicting the difference (error) in measured time, as a function of the average time, for each stride as derived from the app and GAITRite mat.

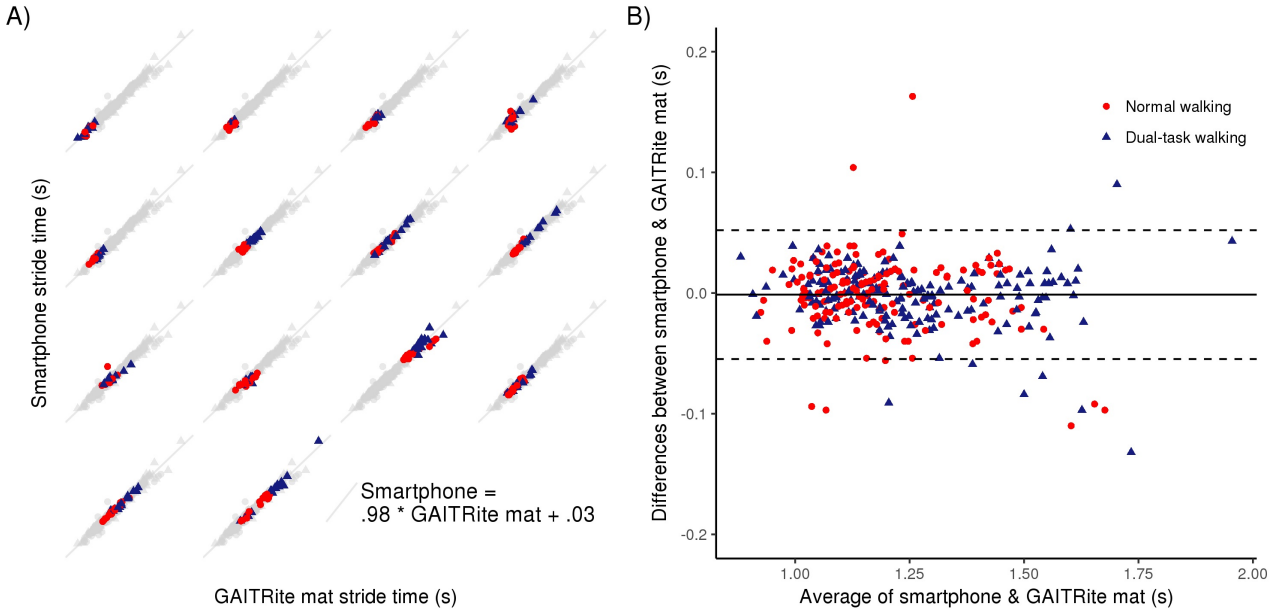


Figure 6. Relationship between dual-task costs to stride time as derived from a smartphone placed in the pocket and from a GAITRite mat. Dual-task cost was defined as the percentage change in average stride time derived from each pair of normal and dual-task walking trials. Dual-task costs as measured by the app and the GAITRite mat were strongly correlated with one another (r^2=.95, P<.001). The orthogonal best fit line of these data had a slope of approximately 1 and an intercept of 0.

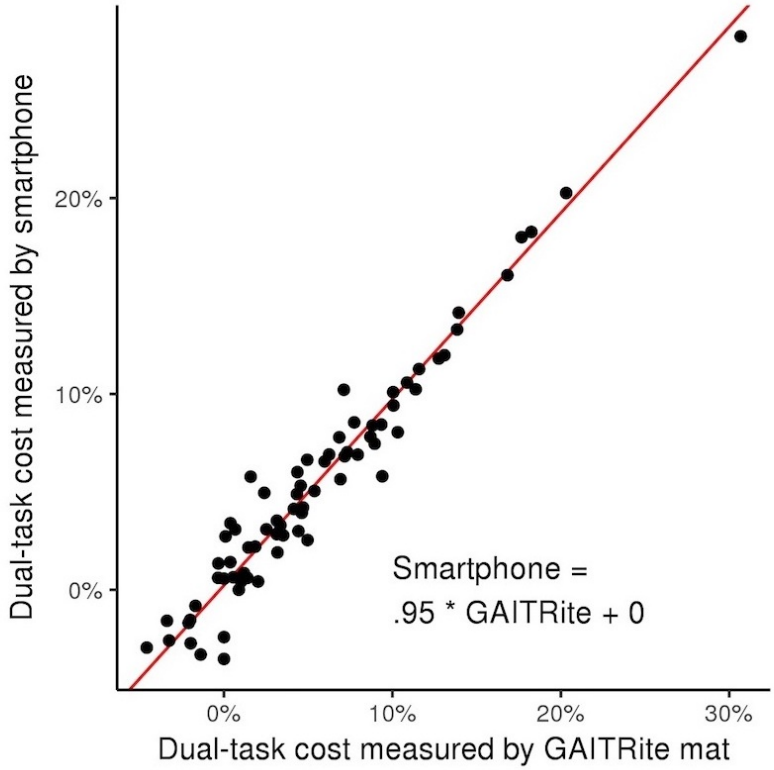


Table 1. Test-retest reliability of laboratory- and home-based assessments of average stride times during normal and dual-task walking.

Tests	Normal walking			Dual-task walking		
	ICC[a]	*P* value	95% CI	ICC	*P* value	95% CI
GAITRite mat						
Within-visit	.94	<.001	.88-.98	.89	<.001	.77-.95
Between-visit	.93	<.001	.79-.97	.83	<.001	.52-.94
Smartphone app						
Laboratory assessment						
Within-visit	.94	<.001	.88-.98	.90	<.001	.79-.96
Between-visit	.92	<.001	.77-.97	.83	<.001	.51-.94
Home assessment	.83	<.001	.62-.94	.82	<.001	.60-.94
Laboratory vs home	.87	<.001	.57-.96	.89	<.001	.65-.97

[a]ICC: intraclass correlation coefficient.

Linear regression models further indicated that this correlation was unaffected by task condition, laboratory visit number, or self-report of pocket tightness.

The dual-task costs to average stride time derived from the app and the GAITRite mat were also highly correlated (P<.001, r^2=.95; Figure 6). This correlation was also independent of laboratory visit number and self-reported pocket tightness.

The average magnitude of error of individual app-derived stride times, as compared with the corresponding stride time derived from the gait mat, was 16.9 (SD 9.0) ms. Figure 5 (part B) depicts a Bland-Altman plot, which illustrates that the magnitude of error was not noticeably influenced by stride time. ANOVA models further indicated that the magnitude of error was similar across laboratory visits (F_{1346}=0.24, P=.63) and was unaffected by either walking conditions (F_{1346}=0.03, P=.86) or pocket tightness (F_{2346}=0.91, P=.40).

Reliability of Smartphone-Measured Stride Time

Average stride times derived from each app trial—for both normal and dual-task walking—demonstrated excellent test-retest reliability across repeated trials within laboratory assessments, across trials between the 2 laboratory assessments separated by 1 week, and across home assessment days (Table 1). In general, we observed that ICC values were (1) slightly higher for trials of normal walking than for dual-task walking, (2) similar in value between app- and GAITRite-based measurements within each laboratory visit and between 2 laboratory visits, and (3) similar in value for home assessments and for laboratory assessments.

Effects of Participant Characteristics, Walking Condition, and Setting on Stride Time

Average stride times were not significantly correlated with participant height or body mass. Stride times were longer (F_{1158}=4.67, P=.03) when dual tasking (mean 1.18, SD 0.16 s) than when walking normally (mean 1.05, SD 0.16 s). Testing setting (ie, laboratory vs home) did not affect average stride times (F_{1158}=0.001, P=.99).

Discussion

This study provides a proof-of-concept in healthy adults that a smartphone placed in the front pocket of one's pants or shorts can provide multimedia instructions to the participant and accurately measure stride times during walking under different experimental conditions. The app can detect major turns and compute average stride times during forward walking with high test-retest reliability within a laboratory or home setting.

Body-worn sensors, including those contained within smartphones, can be used to capture the kinematic properties of gait. Previous work has typically secured the smartphone or sensor tightly to the individual's trunk [11,22-24] or lower extremities [10,25,26]. While that approach has been proven to enable measurement of gait metrics with enough sensitivity to distinguish between disease states, it has used additional equipment (eg, Velcro or elastic straps) together with trained personnel in a laboratory setting to provide assessment instructions. Our approach places the phone in the pocket, provides automated instructions to the participant, and uploads acquired data automatically via Wi–Fi to cloud-based storage, thus providing a widely accessible and cost-effective tool for the assessment of walking within both laboratory and nonlaboratory settings. Such a tool may be particularly useful and cost effective for large-scale national or international studies of mobility by obviating the need for local research staff to instruct participants or apply special instruments.

Walking in everyday life is frequently conducted while executing cognitive tasks. Serial subtraction is most typically used within laboratory dual-task paradigms because it is easily implemented, disrupts the gait of even healthy adults (see Figure 5, part A), and produces measurable dual-task costs that are sensitive to concussion [27,28], aging [20,29], future falls [30,31], and cognitive decline [32,33]. Dual-task assessments are influenced by the instructions provided to the participant prior to each trial of walking, especially with respect to task prioritization [34,35]. In this study, use of the smartphone to provide standardized instructions via a combination of animated, written, and audible instructions led to excellent test-retest

reliability of average stride times derived from dual-task trials, both within the laboratory and at home, that were comparable with or even higher than published reports of similar assessments led by trained personnel [23,36]. Future work is therefore warranted to test and optimize this smartphone approach in older adults and those with varying levels of cognitive and physical impairment. Moreover, efforts are needed to use smartphone voice recognition software to quantify serial-subtraction performance in order to adjust dual-task cost outcomes for this important variable and to standardize cognitive-task difficulty across individuals.

The proposed method of identifying periods of walking that likely included turning, and subsequently removing these periods from the calculation of stride times, led to excellent test-retest reliability of average stride time. It is expected that the future development and application of more sophisticated turn identification algorithms will further improve test-retest reliability of this and other metrics by ensuring that strides included in subsequent analyses were not influenced by turning. Moreover, turning is critical to the navigation of one's environment, and the kinematic characteristics of turning indeed provide important insight into the integrity of the locomotor control system [37-39]. Such an approach that leads to the accurate detection of a turn during remote walking assessments would also be highly valuable by enabling assessment of the kinematic properties of the turn itself.

This study has provided evidence that stride timing—a clinically meaningful outcome of locomotor control—can be accurately and reliably derived from kinematic data acquired by the smartphone when placed in the pocket when walking. Future work is warranted to establish the capability of this app to derive other clinically meaningful metrics of locomotor control, such as walking speed, swing and stance timing, or stride-time variability [17,40,41]. Finally, this study was focused on demonstrating the feasibility of assessing gait remotely using a smartphone and was thus completed in a relatively small cohort of healthy young adults. Larger studies are now needed to establish the validity and reliability of this method in more heterogeneous populations and in those with abnormal gait patterns.

Acknowledgments

This study was funded in part by the Football Players Health Study at Harvard University. BM was supported by the US National Institute on Aging (1-K01-AG044543–01A1), the Boston Claude D. Pepper Older Americans Independence Center (P30-AG013679), and a pilot grant from the Marcus Applebaum Foundation at the Institute for Aging Research. LL holds the Irving and Edyth S. Usen and Family Chair in Geriatric Medicine at Hebrew SeniorLife. APL was supported in part by the Sidney R. Baer Jr. Foundation, the US National Institutes of Health (R01 MH100186, R01 HD069776, R01 NS073601, R21 MH099196, R21 NS085491, R21 NS082870, R21 HD07616), and Harvard Catalyst | The Harvard Clinical and Translational Science Center (NCRR and the NCATS NIH, UL1 RR025758). The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic health care centers, the National Institutes of Health, the National Football League Players Association, or the Sidney R. Baer Jr. Foundation.

Conflicts of Interest

APL serves on the scientific advisory boards for Magstim, Nexstim, Neuronix, Starlab Neuroscience, Neuroelectrics, Axilum Robotics, Constant Therapy, and Neosync, and is listed as an inventor on several issued and pending patents on the real-time integration of transcranial magnetic stimulation with electroencephalography and magnetic resonance imaging.

References

1. Smith E, Cusack T, Cunningham C, Blake C. The influence of a cognitive dual task on the gait parameters of healthy older adults: a systematic review and meta-analysis. J Aging Phys Act 2017 Oct 01;25(4):671-686. [doi: 10.1123/japa.2016-0265] [Medline: 28253049]
2. Barak Y, Wagenaar RC, Holt KG. Gait characteristics of elderly people with a history of falls: a dynamic approach. Phys Ther 2006 Nov;86(11):1501-1510. [doi: 10.2522/ptj.20050387] [Medline: 17079750]
3. Verghese J, Holtzer R, Lipton RB, Wang C. Quantitative gait markers and incident fall risk in older adults. J Gerontol A Biol Sci Med Sci 2009 Aug;64(8):896-901 [FREE Full text] [doi: 10.1093/gerona/glp033] [Medline: 19349593]
4. Verghese J, Lipton RB, Hall CB, Kuslansky G, Katz MJ, Buschke H. Abnormality of gait as a predictor of non-Alzheimer's dementia. N Engl J Med 2002 Nov 28;347(22):1761-1768. [doi: 10.1056/NEJMoa020441] [Medline: 12456852]
5. Beauchet O, Annweiler C, Callisaya ML, De Cock AM, Helbostad JL, Kressig RW, et al. Poor gait performance and prediction of dementia: results from a meta-analysis. J Am Med Dir Assoc 2016 Jun 01;17(6):482-490 [FREE Full text] [doi: 10.1016/j.jamda.2015.12.092] [Medline: 26852960]
6. Ebersbach G, Baas H, Csoti I, Müngersdorf M, Deuschl G. Scales in Parkinson's disease. J Neurol 2006 Aug;253 Suppl 4:IV32-IV35. [doi: 10.1007/s00415-006-4008-0] [Medline: 16944355]
7. Lord SE, Rochester L. Measurement of community ambulation after stroke: current status and future developments. Stroke 2005 Jul;36(7):1457-1461 [FREE Full text] [doi: 10.1161/01.STR.0000170698.20376.2e] [Medline: 15947265]
8. Kaufman M, Moyer D, Norton J. The significant change for the timed 25-foot walk in the multiple sclerosis functional composite. Mult Scler 2000 Aug;6(4):286-290. [doi: 10.1177/135245850000600411] [Medline: 10962550]

9. Haaxma CA, Bloem BR, Borm GF, Horstink MWIM. Comparison of a timed motor test battery to the Unified Parkinson's Disease Rating Scale-III in Parkinson's disease. Mov Disord 2008 Sep 15;23(12):1707-1717. [doi: 10.1002/mds.22197] [Medline: 18649395]
10. Pepa L, Verdini F, Spalazzi L. Gait parameter and event estimation using smartphones. Gait Posture 2017 Sep;57:217-223. [doi: 10.1016/j.gaitpost.2017.06.011] [Medline: 28667903]
11. Ellis RJ, Ng YS, Zhu S, Tan DM, Anderson B, Schlaug G, et al. A validated smartphone-based assessment of gait and gait variability in Parkinson's disease. PLoS One 2015;10(10):e0141694 [FREE Full text] [doi: 10.1371/journal.pone.0141694] [Medline: 26517720]
12. Capecci M, Pepa L, Verdini F, Ceravolo MG. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. Gait Posture 2016 Oct;50:28-33. [doi: 10.1016/j.gaitpost.2016.08.018] [Medline: 27567449]
13. Bilney B, Morris M, Webster K. Concurrent related validity of the GAITRite walkway system for quantification of the spatial and temporal parameters of gait. Gait Posture 2003 Feb;17(1):68-74. [Medline: 12535728]
14. Hausdorff JM, Rios DA, Edelberg HK. Gait variability and fall risk in community-living older adults: a 1-year prospective study. Arch Phys Med Rehabil 2001 Aug;82(8):1050-1056. [doi: 10.1053/apmr.2001.24893] [Medline: 11494184]
15. Hausdorff JM, Cudkowicz ME, Firtion R, Wei JY, Goldberger AL. Gait variability and basal ganglia disorders: stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. Mov Disord 1998 May;13(3):428-437. [doi: 10.1002/mds.870130310] [Medline: 9613733]
16. Montero-Odasso M, Muir SW, Speechley M. Dual-task complexity affects gait in people with mild cognitive impairment: the interplay between gait variability, dual tasking, and risk of falls. Arch Phys Med Rehabil 2012 Feb;93(2):293-299. [doi: 10.1016/j.apmr.2011.08.026] [Medline: 22289240]
17. Hausdorff JM. Gait dynamics, fractals and falls: finding meaning in the stride-to-stride fluctuations of human walking. Hum Mov Sci 2007 Aug;26(4):555-589 [FREE Full text] [doi: 10.1016/j.humov.2007.05.003] [Medline: 17618701]
18. Strike SC, Taylor MJD. The temporal-spatial and ground reaction impulses of turning gait: is turning symmetrical? Gait Posture 2009 Jun;29(4):597-602. [doi: 10.1016/j.gaitpost.2008.12.015] [Medline: 19195890]
19. Yogev-Seligmann G, Hausdorff JM, Giladi N. The role of executive function and attention in gait. Mov Disord 2008 Feb 15;23(3):329-42; quiz 472 [FREE Full text] [doi: 10.1002/mds.21720] [Medline: 18058946]
20. Beurskens R, Bock O. Age-related deficits of dual-task walking: a review. Neural Plast 2012;2012:131608 [FREE Full text] [doi: 10.1155/2012/131608] [Medline: 22848845]
21. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. J Clin Chem Clin Biochem 1983 Nov;21(11):709-720. [Medline: 6655447]
22. González RC, López AM, Rodriguez-Uría J, Alvarez D, Alvarez JC. Real-time gait event detection for normal subjects from lower trunk accelerations. Gait Posture 2010 Mar;31(3):322-325. [doi: 10.1016/j.gaitpost.2009.11.014] [Medline: 20034797]
23. Godfrey A, Del Din S, Barry G, Mathers JC, Rochester L. Instrumenting gait with an accelerometer: a system and algorithm examination. Med Eng Phys 2015 Apr;37(4):400-407 [FREE Full text] [doi: 10.1016/j.medengphy.2015.02.003] [Medline: 25749552]
24. Nishiguchi S, Yamada M, Nagai K, Mori S, Kajiwara Y, Sonoda T, et al. Reliability and validity of gait analysis by android-based smartphone. Telemed J E Health 2012 May;18(4):292-296. [doi: 10.1089/tmj.2011.0132] [Medline: 22400972]
25. Tao W, Liu T, Zheng R, Feng H. Gait analysis using wearable sensors. Sensors (Basel) 2012;12(2):2255-2283 [FREE Full text] [doi: 10.3390/s120202255] [Medline: 22438763]
26. Takeda R, Tadano S, Todoh M, Morikawa M, Nakayasu M, Yoshinari S. Gait analysis using gravitational acceleration measured by wearable sensors. J Biomech 2009 Feb 09;42(3):223-233. [doi: 10.1016/j.jbiomech.2008.10.027] [Medline: 19121522]
27. Lee H, Sullivan SJ, Schneiders AG. The use of the dual-task paradigm in detecting gait performance deficits following a sports-related concussion: a systematic review and meta-analysis. J Sci Med Sport 2013 Jan;16(1):2-7. [doi: 10.1016/j.jsams.2012.03.013] [Medline: 22609052]
28. Howell DR, Osternig LR, Chou L. Dual-task effect on gait balance control in adolescents with concussion. Arch Phys Med Rehabil 2013 Aug;94(8):1513-1520. [doi: 10.1016/j.apmr.2013.04.015] [Medline: 23643687]
29. Hausdorff JM, Schweiger A, Herman T, Yogev-Seligmann G, Giladi N. Dual-task decrements in gait: contributing factors among healthy older adults. J Gerontol A Biol Sci Med Sci 2008 Dec;63(12):1335-1343 [FREE Full text] [Medline: 19126846]
30. Springer S, Giladi N, Peretz C, Yogev G, Simon ES, Hausdorff JM. Dual-tasking effects on gait variability: the role of aging, falls, and executive function. Mov Disord 2006 Jul;21(7):950-957. [doi: 10.1002/mds.20848] [Medline: 16541455]
31. Herman T, Mirelman A, Giladi N, Schweiger A, Hausdorff JM. Executive control deficits as a prodrome to falls in healthy older adults: a prospective study linking thinking, walking, and falling. J Gerontol A Biol Sci Med Sci 2010 Oct;65(10):1086-1092 [FREE Full text] [doi: 10.1093/gerona/glq077] [Medline: 20484336]

32. Montero-Odasso MM, Sarquis-Adamson Y, Speechley M, Borrie MJ, Hachinski VC, Wells J, et al. Association of dual-task gait with incident dementia in mild cognitive impairment: results from the Gait and Brain Study. JAMA Neurol 2017 Jul 01;74(7):857-865. [doi: 10.1001/jamaneurol.2017.0643] [Medline: 28505243]
33. Muir SW, Speechley M, Wells J, Borrie M, Gopaul K, Montero-Odasso M. Gait assessment in mild cognitive impairment and Alzheimer's disease: the effect of dual-task challenges across the cognitive spectrum. Gait Posture 2012 Jan;35(1):96-100. [doi: 10.1016/j.gaitpost.2011.08.014] [Medline: 21940172]
34. Kelly VE, Janke AA, Shumway-Cook A. Effects of instructed focus and task difficulty on concurrent walking and cognitive task performance in healthy young adults. Exp Brain Res 2010 Nov;207(1-2):65-73 [FREE Full text] [doi: 10.1007/s00221-010-2429-6] [Medline: 20931180]
35. Yogev-Seligmann G, Rotem-Galili Y, Mirelman A, Dickstein R, Giladi N, Hausdorff JM. How does explicit prioritization alter walking during dual-task performance? Effects of age and sex on gait speed and variability. Phys Ther 2010 Feb;90(2):177-186 [FREE Full text] [doi: 10.2522/ptj.20090043] [Medline: 20023000]
36. Byun S, Han JW, Kim TH, Kim KW. Test-retest reliability and concurrent validity of a single tri-axial accelerometer-based gait analysis in older adults with normal cognition. PLoS One 2016;11(7):e0158956 [FREE Full text] [doi: 10.1371/journal.pone.0158956] [Medline: 27427965]
37. Patla AE, Adkin A, Ballard T. Online steering: coordination and control of body center of mass, head and body reorientation. Exp Brain Res 1999 Dec;129(4):629-634. [Medline: 10638436]
38. Mancini M, El-Gohary M, Pearson S, McNames J, Schlueter H, Nutt JG, et al. Continuous monitoring of turning in Parkinson's disease: rehabilitation potential. NeuroRehabilitation 2015;37(1):3-10 [FREE Full text] [doi: 10.3233/NRE-151236] [Medline: 26409689]
39. Salarian A, Zampieri C, Horak FB, Carlson-Kuhta P, Nutt JG, Aminian K. Analyzing 180 degrees turns using an inertial system reveals early signs of progression of Parkinson's disease. Conf Proc IEEE Eng Med Biol Soc 2009;2009:224-227 [FREE Full text] [doi: 10.1109/IEMBS.2009.5333970] [Medline: 19964471]
40. Hausdorff JM. Gait variability: methods, modeling and meaning. J Neuroeng Rehabil 2005 Jul 20;2:19 [FREE Full text] [doi: 10.1186/1743-0003-2-19] [Medline: 16033650]
41. Hausdorff JM. Stride variability: beyond length and frequency. Gait Posture 2004 Dec;20(3):304; author reply 305. [doi: 10.1016/j.gaitpost.2003.08.002] [Medline: 15531178]

Abbreviations

AUC: area under the curve
ICC: intraclass correlation coefficient

Edited by G Eysenbach; submitted 23.08.17; peer-reviewed by G Barry, K Theunissen; comments to author 15.09.17; revised version received 24.11.17; accepted 24.11.17; published 30.01.18.

Please cite as:
Manor B, Yu W, Zhu H, Harrison R, Lo OY, Lipsitz L, Travison T, Pascual-Leone A, Zhou J
Smartphone App–Based Assessment of Gait During Normal and Dual-Task Walking: Demonstration of Validity and Reliability
JMIR Mhealth Uhealth 2018;6(1):e36
URL: http://mhealth.jmir.org/2018/1/e36/
doi:10.2196/mhealth.8815
PMID:29382625

Original Paper

A Breastfeed-Promoting Mobile App Intervention: Usability and Usefulness Study

Chih-Jau Wang[1*], BSc; Pimwadee Chaovalit[1*], PhD; Suporn Pongnumkul[1*], PhD

National Electronics and Computer Technology Center, Pathum Thani, Thailand

[*]all authors contributed equally

Corresponding Author:
Chih-Jau Wang, BSc
National Electronics and Computer Technology Center
112 Phahonyothin Road, Khlong Nueng, Khlong Luang District
Pathum Thani, 12120
Thailand
Phone: 66 2 564 6900 ext 72293
Email: wang.chihjau@gmail.com

Abstract

Background: Breastfeeding is proven to have lasting health benefits for both mothers and infants; however, 6-month exclusive breastfeeding rate remains below 20% in Thailand. Although the number of research literature and commercial apps for breastfeeding women is significantly growing, they are country-specific and restricted to English-speaking users. There exists a major knowledge gap on how mobile health apps could support breastfeeding in Thailand. To address these gaps, MoomMae has been developed with the intention to support Thai women in breastfeeding outside of their homes and in keeping their feeding records.

Objective: The aim of this study was to evaluate the usability and usefulness of MoomMae, a mobile phone app designed to support breastfeeding women.

Methods: Our study was reviewed and approved by Thailand's National Science and Technology Development Agency (NSTDA) ethics committee. A total of 21 breastfeeding women with at least one Android phone or tablet were recruited via convenience and snowball sampling. The study process for each participant was as follows: the participant was requested to attend a preuse interview and given the app to use for 4 weeks. Following this period, a postuse interview was conducted to examine the usability and usefulness of the app. Both sessions were held individually and audiorecorded for qualitative analysis.

Results: The mean scores of usability and usefulness from the postuse survey were 4.33 (SD 0.87; range 1-5) and 4.60 (SD 0.74; range 2-5). Our qualitative analysis revealed a total of 137 feedbacks: 71 related to usability and 66 associated with usefulness. A further sentimental analysis showed that comments on usability were generally negative (59 negative, 11 positive, and 1 neutral), and comments on usefulness were relatively positive (56 positive, 9 negative, and 1 neutral). We discovered 26 unique design issues and proposed recommendations for future improvement.

Conclusions: Our usability and usefulness assessment of MoomMae demonstrated that MoomMae has a great potential to be a useful self-management tool for breastfeeding mothers in Thailand. The qualitative analysis suggested that the app is supportive of breastfeeding on demand, but the flow and inputs of the app should be redesigned to be more intuitive. For future implementations, the most desirable feature is a pump-reminding notification system.

(JMIR Mhealth Uhealth 2018;6(1):e27) doi:10.2196/mhealth.8337

KEYWORDS

mobile health; breast feeding; mobile applications; health promotion; usability; usefulness

Introduction

Breastfeeding

Breast milk is universally recognized as the best food for newborns. Studies have scientifically shown that breastfeeding provides optimal nutrients for infants, strengthens their immune system, and improves mother-and-child bonding [1,2]. Breastfeeding women also have lower risks of breast and ovarian cancer [3]. The World Health Organization (WHO), therefore, recommends exclusive breastfeeding for up to 6 months and continued breastfeeding for 1 to 2 years [4]. Despite the documented benefits, committing to breastfeeding can be challenging for mothers. As a result, exclusive breastfeeding

rates at the recommended 6 months are under 39% at the global scale [5]. In Thailand, with an average of 715,000 newborns yearly between 2010 and 2015, the initiation rates of breastfeeding was 46% and the exclusive breastfeeding rates up to 6 months was only 12% [5].

One leading factor for discontinuation of breastfeeding is the discomfort to breastfeed or breast pump in the public [6,7]. Breastfeeding women described that they are judged or stopped from breastfeeding in the public as societies expect them to do it at home [8]. At the same time, mothers are expected to leave home for work, do grocery shopping, and take their babies to hospitals. Mothers often end up breastfeeding or breast pumping in public toilets or in personal cars, and some choose to give formula milk instead because of inconvenience [9].

Following a baby's feeding pattern is another cumbersome task for mothers, as the quantity, duration, and frequency of feeding varies widely between mother-baby pairs and over time [10]. Some infants may spend 5 min on each breast and others can take up to 30 min per breast. Keeping a feeding record, therefore, helps mothers to learn when their child typically gets hungry and to know if their milk supply is sufficient. Sufficient milk supply consequently helps to build up the mothers' confidence in their ability to breastfeed [11]. The feeding log also helps doctors to reassure that breastfeeding is going well and that the baby's weight gain is on standard [12]. For instance, feedings that last more than 40 min repeatedly might indicate a sign of low milk supply or a baby's wrong sucking skills.

Information Technology for Breastfeeding Women

Research has shown that technological interventions were more effective in promoting breastfeeding when compared with traditional face-to-face interventions [13]. These technologies may vary in forms from message prompts [14,15], multimedia files [16,17], computer programs [18,19], to online websites [20-24]. In addition to this, mobile apps have become a popular platform used to support breastfeeding [25].

In the market, there exists a number of apps for breastfeeding mothers. Mobile apps for tracking feeding and pumping logs are vastly available. For example, Breastfeeding Tracker Pumping allows mothers to track the time and duration of different types of feeding [26]. Other apps come with the features of recording the baby's feeds, sleep sessions, diaper changes, and growth [27-31]. Mobile apps for finding public breastfeeding places include MomsPumpHere, Mam Lactation, LatchMe in the United States [32-34], FeedFinder in the United Kingdom [35], Baby Places in Europe [36], MamaMap in Switzerland [37], and Nursing Room SG in Singapore [38]. However, these apps are specific to areas and countries. To the best of our knowledge, there is no such app in Thailand. Information regarding public breastfeeding rooms in Thailand are indexed online [39] and is shared via blogging such as Little Panda's Travel Diary and Khajochi Blog [40,41].

Usability of Mobile Health App

The advances in mobile technologies have made mobile phone apps more effective to promote health behavior change [42,43]. The ubiquitous nature of mobile phones has also promised greater patient engagement. However, its usability is often challenged by the hardware limitations, including small screen and limited input capabilities [44]. Studies have shown that mobile health (mHealth) apps are not always effective, and there is room for improvement [45,46]. New platforms should ideally be assessed to ensure usability.

Usability evaluation is the key to enhance acceptability and can be examined by several direct and indirect methods [47]. Two most common direct testing methods are thinking aloud and performance measurement. The thinking-aloud method is used to gain insights into real feelings of the subjects by asking them to navigate through the system and ask them to voice their thoughts aloud. In the performance measurement method, subjects are asked to perform a set of predefined tasks while investigators make some quantifiable measurements—such as completion time for each task. The performance measurement method is suitable for analyzing the usability of two competing approaches or designs. Indirect testing methods include questionnaire and interview. These methods are useful to assess parameters that are difficult to measure objectively, such as satisfaction and frustration.

Besides the direct and indirect protocols, heuristic evaluation has been applied to enhance user experience by identifying usability problems of a user interface (UI) that violate the design principles. The original heuristics was developed by Jakob Nielsen in the 1990s [48] with 10 heuristics to evaluate UIs on desktop devices. This set of heuristics, although widely used today, faces challenges to evaluate other devices. The research community has been putting efforts to modify these heuristics to be more suitable for mobile computing. In 2006, a group of researchers reported a new set of heuristics for mobile apps [49], which has been used in several studies [44,50,51].

Objectives

Although the number of research literature and commercial apps for breastfeeding women is vastly expanding, these research and apps are restricted to English-speaking users. Existing apps for finding public breastfeeding places are also specific to countries. As a result, there is a concerning lack of the knowledge on how mHealth apps could support breastfeeding in Thailand. We address these gaps by evaluating usability and usefulness of MoomMae, a mobile app that aims to support women in Thailand to breastfeed comfortably in the public and keep their feeding and pumping logs efficiently.

Methods

Recruitment

A convenience and snowball sample of 21 breastfeeding women was enrolled in this study. Participants were approached via recruitment posters through two channels: 12 participants were members of a Facebook group named *Nom Mae Happy* (meaning happy breast milk) and 9 participants worked at National Science and Technology Development Agency (NSTDA), Thailand. The posters included the app introduction, inclusion criteria, and registration link. After registration, participants were contacted by telephone to give a brief introduction to the study and make an appointment for the subsequent interviews upon agreement with the subjects. Our

inclusion criteria were as follows: subjects must be over 20 years old, own an Android mobile phone or tablet, live in Bangkok area, and be able to attend two 1-hour interview sessions. Participants were excluded if they were pregnant or were not competent to provide feedback by themselves.

Ethical Review

This study was approved by the NSTDA ethics board of committee. The contents of participant information, informed consent form, recruiting posters, and registration form were reviewed. All participants provided full written informed consent before the interview sessions.

Measurement and Analysis

Figure 1 summarizes the evaluation flow of our preuse and postuse structured interviews. Both sessions were audio-recorded, held individually, and led by the author who was not involved in the design and development of the app. The preuse sessions were held at the beginning of the 4-week trial period. On the day, a structured interview was performed to gather participant demographics, their current feeding options, prior experience with mobile phones, and expectation toward the app. After the 4-week app usage, the postuse sessions were conducted and consisted of two parts: a survey and a structured interview.

The postuse survey asked participants to rate usability and usefulness of the app based on the Likert scale. Possible scores were 1, 2, 3, 4, and 5, ranging from strongly disagree to strongly agree. Questions were adapted from a previously published study [52]. The usability questions compose of five attributes: learnability, efficiency, memorability, errors, and satisfaction. The usefulness questions were based on three criteria: app features, comfort after use, and intention to use. Results were analyzed to report means and standard deviations (SDs) of each component.

The postuse interviews were structurally performed and asked participants to share their insight into the app usage. Data collected through hand-written notes and recorded spoken responses were qualitatively analyzed. Each response was mapped to one app feature and labeled as positive, neutral, or negative. Results on usability were further analyzed using heuristic evaluation to identify design issues and to provide design recommendation. Results on usefulness were matched to the “Ten Steps to Successful Breastfeeding”, hereafter referred to as “The Ten Steps” [53], to reveal our roles in supporting breastfeeding and suggestions for future implementation.

App Description

MoomMae is an mHealth app designed to serve as a supportive health technology for mothers to achieve their breastfeeding goals. The app consists of three main functions related to breastfeeding: feeding record, pumping record, and feeding rooms. In this study, we aim to evaluate the usability and usefulness of these three functions. Screenshots of the three functions in original language are available in Multimedia Appendices 1-3.

In the “feeding record” function (Figure 2; please refer to Multimedia Appendix 1 for the original language), mothers are able to insert new feeding records and view the summary or history of their logs. The main screen shows the daily and overall statistics of the records. Users can swipe left or right to go to the previous or the next child. Mothers can click the “save” button at the top of the screen to make a new feeding data. The “summary” button navigates to the summary page, which shows the graph of the feedings in weekly or monthly mode and in breastfeeding or bottle-feeding mode. The “history” tab leads to the history page, which lists all feeding data.

Figure 1. Evaluation flowchart.

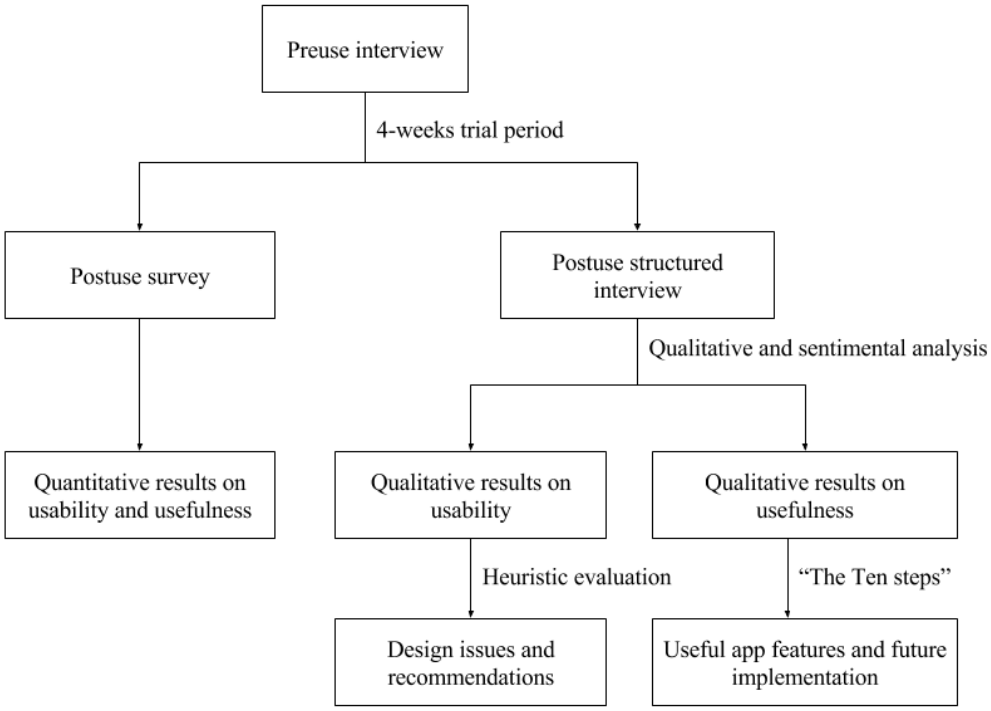


Figure 2. Screenshots of the "feeding record" function. (1) The main screen of the "feeding record" function; (2) Entering a new feeding log; (3) Summary of feeding record; (4) Feeding history. These screenshots are translated into English.

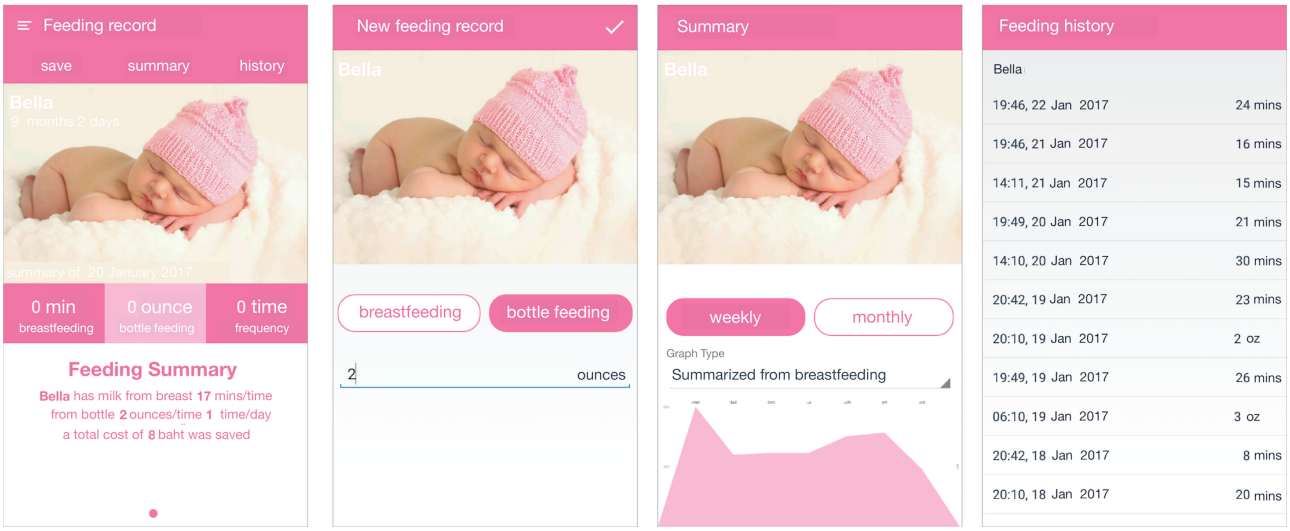
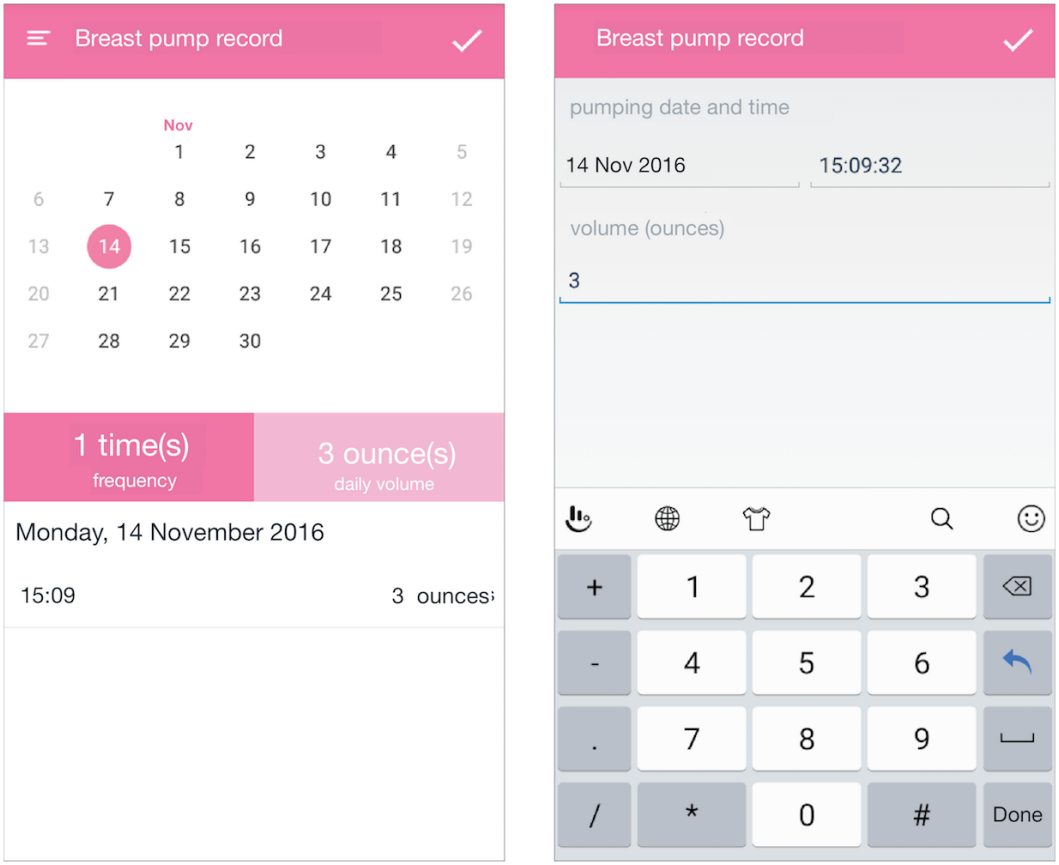


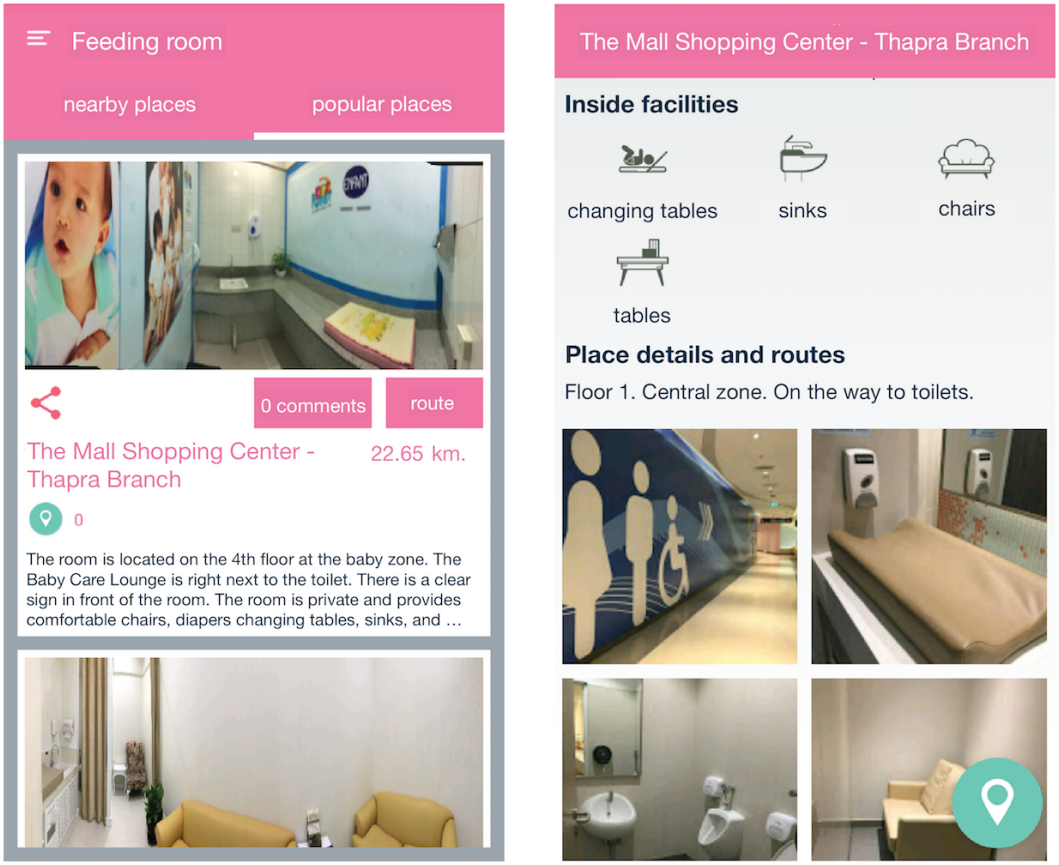
Figure 3. Screenshots of the breast pump function. (1) The breast pump function; (2) Making a new pumping record.



The “pumping record” function (Figure 3; screenshots in original language are available in Multimedia Appendix 2) allows mothers to insert pumping records and see daily statistics of their pumping records. The calendar at the top of the main screen is there for mothers to select a date to view their records or insert a new record. Once a date on the calendar is clicked, the daily statistics and list table view change to display only data of the selected date. To insert a new record, mothers can select a date on the calendar and click the check button at the top-right corner to go to the insert page.

The “feeding rooms” function (Figure 4; screenshots in original language are available in Multimedia Appendix 3) allows mothers to look for public feeding places. Mothers can click on the tabs at the top to change between nearby or popular feeding rooms. Once mothers click on a place on the list, the app will navigate to the detail page of that place that shows descriptions, facilities, routes, and photos of that room.

Figure 4. Screenshots of the breastfeeding rooms function. (1) The breastfeeding rooms function. (2) The detail page of a breastfeeding room.



Results

Participant Characteristics

Table 1 summarizes demographic characteristics of 21 participants included in this study. The average age of the participants was 32.67 years (range 23-41). Almost all participants (20/21, 95%) exclusively breastfed up to 6 months. The remainder reported to breastfeed 98% of the time: giving formula milk a few times when her child was underweight. The leading motivation to breastfeed was to improve immune systems of their babies (20/21; 95%). The top three places to breastfeed were in personal cars (17/21; 81%), sitting areas in public places (13/21; 62%), and restaurants (13/21; 62%). Seven participants (33%, 7/21) used to note their feeding on papers, with one participant (s3) pointing out that it was difficult to calculate the sum and average on a notebook. Ten participants (48%, 10/21) had searched for a breastfeeding room by asking the staff of the building.

App Usage

In the preuse survey, 5 participants reported to spend 1 to 3 hours a day on mobile phones, 9 spent 3 to 5 hours, and 7 spent more than 5 hours. Participants were also asked to rate how much the following information would assist them to achieve their breastfeeding goals. On the scale of 1 to 5, the information about breastfeeding rooms, feeding records, and money saved received mean scores of 4.67 (SD 0.58; range 3-5), 4.38 (SD 0.92; range 2-5), and 4.62 (SD 0.80; range 2-5), respectively.

In the postuse survey, 16 participants (76%, 16/21) stated that they used the app 4 to 7 days a week. The other 5 participants (24%, 5/21) used the app 1 to 3 days a week. Only 4 participants immediately made a record in the app after they breastfed or breast pumped. Nine participants reported using the app during the night after their children fell asleep, stating it was the time they were more available. Six participants used the app in the morning, and 2 used the app during the daytime. The most frequently used functions for each participant were breastfeeding record (12/21, 57%), breast pump record (8/21, 38%), and breastfeeding room (1/21, 5%). The most satisfied functions for each participant were breastfeeding record (7/21, 33%), breast pump record (7/21, 33%), breastfeeding rooms (4/21, 19%), and breastfeeding summary (3/21, 14%).

Quantitative Results

Usability

Usability scores and the complete question list used in postuse surveys are presented in Table 2. The total average score was 34.83 out of the full score of 40. Mean scores of each question were at least 3.60, which was above the midpoint of the scale 1 to 5. Not every question was answered by all participants. For instance, only a few participants used the app to look for public breastfeeding room; hence, Q4 and Q6 were answered by only 12 participants. The mean score for learnability, memorability, errors, efficiency, and satisfaction were at least 3.60 (SD 0.94), 4.76 (SD 0.54), 4.55 (SD 0.69), 4.48 (SD 0.87), and 4.33 (SD 0.73), respectively.

Table 1. Characteristics of the study participants.

Characteristic	Value
Age in years, mean (SD; range)	32.67 (4.50; 23-41)
Education, n (%)	
≤High school diploma	2 (9)
Bachelor's degree	10 (48)
Graduate degree	9 (43)
Employment, n (%)	
Government officer	10 (47)
Housewife	8 (38)
Employee	1 (5)
Self-employed	1 (5)
Student	1 (5)
Exclusive breastfeeding rates up to 6 months, n (%)	
100	20 (95)
60-99	1 (5)
Hours of mobile phone use per day, n (%)	
1-3 hours	5 (24)
3-5 hours	9 (43)
>5 hours	7 (33)

Table 2. Quantitative results of the usability survey.

Question	Usability attribute	n[a]	Mean (SD)[b]
1. You were able to learn and understand the app quickly when you first used it	Learnability	21	3.81 (1.03)
2. You think the other mothers can learn and understand the app quickly.	Learnability	20	3.60 (0.94)
3. If you were to have a period of not using the app, you think that you can easily pick it up again and remember how to use it	Memorability	21	4.76 (0.54)
4. The app accurately shows the information of public breastfeeding rooms	Errors	11	4.55 (0.69)
5. The app accurately shows your feeding history	Errors	20	4.80 (0.41)
6. The app helps you look for breastfeeding rooms faster	Efficiency	12	4.50 (0.80)
7. The app helps you to record your breastfeeding activities faster and more efficiently	Efficiency	21	4.48 (0.87)
8. Overall, you are satisfied with the app	Satisfaction	21	4.33 (0.73)

[a]Number of participants providing a response for each item.

[b]Out of 5. Responses for each item on a scale of 1 (strongly disagree) to 5 (strongly agree). Some participants indicated items were "not applicable" to their experience.

Usefulness

The results for usefulness (Table 3) were slightly higher than the usability scores in all aspects. The total average score was 36.81. The minimum mean score was 4.33, and the maximum mean score of 4.81. The mean score for app features, comfort after use, and intention to use were at least 4.38 (SD 0.81), 4.33 (SD 1.02), and 4.62 (SD 0.80), respectively.

Qualitative Results

Design Issues

The postuse interviews exposed a total of 71 feedbacks on usability. A further sentimental analysis was conducted, and each comment was labeled as positive, neutral, or negative. The results showed that the majority of comments were negative—59 negative, 11 positive, and 1 neutral. Of the negative reports obtained, we revealed 26 unique design issues of the app that could affect the efficiency of use. The problems were then categorized by app features and mapped to one of the eight usability heuristics for mobile devices (Table 4).

The majority of the issues (8 out of 26) were identified to violate the second heuristics (H2)—match between system and the real world. The second most violated heuristics was H5 (6 issues), describing ease of input, screen readability, and glanceability and H6 (6 issues), describing flexibility, efficiency of use, and personalization. Following that, two issues were related to H8—realistic error management. The remaining four issues were associated with H1—visibility of system status and losability or findability of the mobile device, H3—consistency and mapping, H4—good ergonomics and minimalist design, and H7—aesthetic, privacy, and social conventions.

The frequency of these problems ranged from 1 to 6. The most occurring issue (frequency=6) was related to the process of making a new feeding record. The second most common issue (frequency=5) was on the subject of buttons, which do not resemble their actions. The third most expressed concerns (frequency=4) were associated with inputs, breaking down into inputs for volume, remark, and date. Looking at these problems, it is apparent that the app needs to be more intuitive.

Feeding and pumping record. Breaking down into app features, there were three design issues that appeared commonly in both "feeding record" and "pumping record" functions. First, 4 participants struggled to enter volume in ounces in decimal points. They felt that it was user error and they were the only ones who encounter this problem. Second, 4 participants mentioned their desires for writing down textual notes beyond numerical data, such as food they had on that day. Third, 3 participants had concerns with entering multiple records at once. Three participants stated the following:

> *I couldn't enter decimal points, so I had to enter 0.5 ounces less for this time and added that extra 0.5 ounces to the next record. But, that made the data not accurate.* [s15 (participant number); age 31 years; child 9 months]
>
> *Making records on notebooks was more convenient. I could actually write everything I wanted to say about this record, including what I ate on that day.* [s4; age 39 years; child 4 years]
>
> *It took too long to enter a record. As a mother, I had too many responsibilities. I got kids to look after and housework to do. Sometimes, I couldn't make a record right after I fed my baby. I ended up with too many records at the end of the day, but the app didn't support entering multiple records at a time.* [s13; age 28 years; child 5 months]

Feeding Record

For the "feeding record" function, the majority had troubles with knowing where to click to make new (n=6) and past (n=3) feeding records. Another issue related to this function was the absence of saving status after the record has been entered. On the main screen, participants were not able to differentiate between the following items: daily and overall statistics, clickable button, and unclickable plain text. In the summary page, participants experienced difficulties with reading the graph. In the history page, participants struggled with seeking for a particular log as all data were listed on one page without any groupings. Four participants stated the following:

> *When I first entered the main screen of the feeding function, I didn't know where to click to make a new record. The buttons at the top weren't obvious enough.* [s14; age 35 years; child 5 months]
>
> *The app doesn't allow me to enter a record for the past dates, so my data was incomplete.* [s18; age 29 years; child 10 months]
>
> *After I hit the save button, the app didn't state if it was successfully saved or not. The app navigated back to the main screen, but still, I wasn't sure if my records have been made or not.* [s6; age 32 years; child 10 months]
>
> *The feeding history page was really long and difficult to look for a record with a specific date. This made it difficult to change to date and time of the record when I needed to.* [s13; age 28 years; child 5 months]

Table 3. Quantitative results of the usefulness survey.

Question	n[a]	Mean (SD)[b]
You think that the information in #1 to #4 helps you achieve in breastfeeding more:		
1. Information from searching for public breastfeeding rooms	14	4.57 (0.94)
2. Information from recording amounts of both pumped and fed breast milk	17	4.59 (0.62)
3. Information from recording duration of breastfeeding	16	4.38 (0.81)
4. Information about cost you save from breastfeeding	8	4.75 (0.46)
5. You were more comfortable to go out	21	4.33 (1.02)
6. You were more comfortable to breast pump or breastfeed in public areas	21	4.76 (0.44)
7. You plan to continue to use the app	21	4.62 (0.80)
8. You plan to introduce the app to other mothers	21	4.81 (0.51)

[a]Number of participants providing a response for each item.

[b]Out of 5. Responses for each item on a scale of 1 (strongly disagree) to 5 (strongly agree). Some participants indicated items were "not applicable" to their experience.

Table 4. Usability issues resulted from interviews.

App feature and problem	Frequency	Heuristics[a]
Feeding and pumping record		
Inputs for volume (in ounces) do not allow decimal points	4	H2—Match between system and the real world
No free-text remark for feeding and pumping logs	4	H6—Flexibility, efficiency of use, and personalization
Making multiple records at once is not possible	3	H6—Flexibility, efficiency of use, and personalization
Feeding record		
The process of making a new feeding record is not intuitive	6	H2—Match between system and the real world
The process of making past records is not intuitive	3	H6—Flexibility, efficiency of use, and personalization
The graph in summary page is not informative	2	H2—Match between system and the real world
No status shown after saving a record	2	H1—Visibility of system status and losability or findability of the mobile device
The history page is difficult to read	2	H5—Ease of input, screen readability, and glanceability
Daily and overall statistics shown on the main screen are difficult to differentiate	1	H5—Ease of input, screen readability and glanceability
Pumping record		
The process of making a new pumping record is not intuitive	2	H2—Match between system and the real world
Records are not listed in chronological order	1	H2—Match between system and the real world
Feeding rooms		
Searching for a place by its name is not possible	2	H6—Flexibility, efficiency of use, and personalization
Loading takes too long	2	H8—Realistic error management
The lists of nearby and popular places are too informative and difficult to use	1	H4—Good ergonomics and minimalist design
Other		
Inconsistent use of word choices referring to “breast milk”	2	H3—Consistency and mapping
The buttons do not resemble their actions	5	H2—Match between system and the real world
The registration process is not convenient	2	H2—Match between system and the real world
Pressing the back button on the phone forces to exit the app	1	H2—Match between system and the real world
When entering a name, hitting enter on keyboards goes to the next line instead of the next input	1	H2—Match between system and the real world
The app forces to enter personal information in order to use other features	1	H6—Flexibility, efficiency of use, and personalization
Date and time inputs are difficult to use	4	H5—Ease of input, screen readability, and glanceability
Profile pictures are difficult to set up	2	H5—Ease of input, screen readability, and glanceability
The menu icon is too small	1	H5—Ease of input, screen readability, and glanceability
The Web page that the “breastfeeding information” function links to is not readable on mobile phones	1	H5—Ease of input, screen readability, and glanceability
The app processes too slowly	3	H8—Realistic error management
Some pictures are stretched to fit a fixed ratio	1	H7—Aesthetic, privacy, and social conventions

[a]Heuristics identified by a number preceded by an H (ie, H1: heuristic 1, H2: heuristic 2, etc.)

Pumping Record

The “pumping record” function consisted of two usability concerns. First, participants found the main screen difficult to navigate, and as a result, struggled to find the right place to click to insert a new pumping log. Second, records are sorted by the insert time instead of the pumping time. One participant stated the following:

> *When I first entered the pumping record function, I didn’t know where to click at all. I asked my husband to help but he couldn’t figure it out either.* [s14; age 35 years; child 5 months]

Feeding Rooms

In the “feeding rooms” function, participants experienced difficulties when they were going to a specific place and wanted

to search in advance if there were any feeding rooms in that place. Others responded that the listing page took too long to load and was packed with too much information, making it difficult to read. Two participants stated the following:

> *When I needed to go to a particular place, I wanted to know if there is a feeding room there. However, there wasn't any search boxes to search for a place by name. Thus, I needed to go back to Google.* [s8; age 31 years; child 2 months]

> *The feeding room function requires Internet. That makes the app not so effective when searching for a feeding room outside home. I would use my Wi-Fi at home to look for the place, and once I was there, I needed to ask the staff for more information.* [s2; age 40 years; child 1 year 8 months]

Design Recommendation

Our heuristic evaluation revealed the insights into barriers that prevent participants from undergoing joyful experience. A further analysis was carried out to provide a set of design recommendations that could be applied to similar systems and future versions of the app. The results are categorized by heuristics and presented in Table 5.

To improve user engagement, we recommend improving the flexibility of the following features: feeding record, feeding room, start-up page, and registration. For feeding record, the app should allow users to edit time when entering data, as most participants stated that they could not record right after feeding. We also suggest adding an optional text field for remarks of both feeding and pumping records. Furthermore, developers should implement a status bar indicating the status of saving or navigate the app to the location of the inserted record. For feeding room, we agree with participants that, to enhance efficiency, a search box should be added. In start-up page, designers should avoid forcing users to enter personal information to use other features; otherwise, a skip button should be provided. In registration page, the "I agree to terms" checkbox should be placed before the "register" button.

Displaying a long list of information is challenging and should be given extra attention to. Participants experienced difficulties in reading long lists of feeding history and feeding room exhibited in the app. We recommend that feeding records in the history page should be partitioned into smaller groups, ideally by feeding date. For feeding room list, essential information (eg, title) of each room should be enlarged and insignificant information (eg, number of comments) could be faded out to improve readability. Developers should also avoid loading data at once to enhance better loading time. Other minor advices include listing records in chronological orders and labeling all axes on the graph.

All UI components—such as buttons, input fields, and images—used in the app should be carefully designed. For instance, participants had troubles interpreting the meaning of the icons on the buttons; therefore, we encourage designers to place text labels next to the icons or else use text labels only. Important buttons should also be eye-catching and noticeable at the first glance. For input fields, we promote using the day-month-year date input for birthdays instead of the calendar-style date input. Developers should also implement input fields to support real-life conventions such as enabling decimal points for inputs with ounces as unit and replacing the "return" button on the keyboards with "next" when several fields are required. For images, the app should not stretch them to fit the screen width but display them in their original ratios. We recommend to avoid placing text on images and to replace missing ones by default pictures.

Usefulness of System Features

A total of 66 reviews on usefulness were uncovered from the postuse interviews. The obtained comments were grouped into 56 positive and 9 negative comments. When classified by the app features, the majority (36/66) were associated with the "feeding record" and "pumping record" functions. The remaining comments were expressed toward the "feeding rooms" function (19/66) and overall usefulness of the app (11/66).

Feeding and Pumping Records

For the "feeding record" and "pumping record" functions, there were 36 feedbacks captured and were sentimentally classified into 33 positive and 3 negative comments. The negative reviews were related specifically to the graph shown in the summary page, indicating that a more appropriate data visualization technique could be used. From the positive reviews, we discovered three benefits that the app provides: confidence building, feeding-and-pumping volume control, and time management. Participants reported that the app reduced their burdens of both insufficient and excess milk supply. Keeping a regular feeding and pumping records helped them estimate their children's milk and manage their own milk supply more effectively. Additionally, continuation of recording helped mothers develop a more accurate feeding schedule and, consequently, enhanced their personal time management. Examples of participants' feedbacks associated with the three themes mentioned above are illustrated in the following quotes:

> *The part of the app that show how much I saved was really pleasing. I gave my child the best food possible with the lowest price. It made me proud as a mother.* [s14; age 35 years; child 5 months]

> *Keeping a feeding record helps me notice the feeding time. In the past, my mom fed my kid every time he cried. She thought he was hungry, but that wasn't always the case. Now, I know his time. I won't overfeed him and his weight won't exceed the standard.* [s14; age 35 years; child 5 months]

> *My kid was sick, and my doctor kept asking if he ate less than he usually did. I didn't think he could eat less. I thought he ate normally. Now that I start using the app, I can clearly see that his eating trend is dropping.* [s13; age 28 years; child 5 months]

Table 5. The design recommendations.

Heuristics[a]	Design recommendation
H1—Visibility of system status and losability or findability of the mobile device	Implement a status bar or navigate the app to the location of the inserted record
H2—Match between system and the real world	• Make sure the add button is obvious and noticeable at the first glance • List all records by chronological order • Enable decimal points for ounces input fields • Make sure icons on buttons resemble their actions. Include text on buttons if appropriate • Specify the input method type to replace the return button on keyboards with done or next • Label the graph axes
H3—Consistency and mapping	Use the word "breast milk" instead of "milk" alone. If appropriate to the context, apply it across the app
H4—Good ergonomics and minimalist design	For each item on the feeding rooms list, enlarge important information (such as title) and fade out insignificant information (such as the comment count)
H5—Ease of input, screen readability, and glanceability	• In feeding history, group feeding records by dates and feeding method • Avoid putting text on images • In profile setup page, omit the calendar input. Use regular day/month/year date input • Make sure important buttons are eye-catching and noticeable at the first glance
H6—Flexibility, efficiency of use, and personalization	• Add a free text field for remarks of each feeding and pumping record • Redesign the process of making new feeding records. Allow users to edit time while entering data • Add a search box at the top of the feeding room list • Avoid forcing users to enter personal information or provide a skip button • Place the "I agree to terms" checkbox before the "register" button
H7—Aesthetic, privacy, and social conventions	Avoid stretching images to fit the screen size. Replace missing image with appropriate default image
H8—Realistic error management	In feeding rooms list, avoid loading all information at once

[a]Heuristics identified by a number preceded by an H (ie, H1: heuristic 1, H2: heuristic 2, etc.)

Feeding Room

A total of 14 positive and 5 negative comments were expressed toward the usefulness of the "feeding room" feature. The negative feedbacks are related to the limited number of feeding rooms exhibited in the app, especially to the area outside of downtown Bangkok. Of the positive comments, participants indicated that the information about feeding rooms made breastfeeding in the public more effective, pleasant, and comfortable, as illustrated in the following quotes:

> *Previously, I needed to carry a piece of cloth with me everywhere to cover my breast. Now, I can look or feeding places in advance. I can manage myself better. I knew where and when to feed my kids. I knew what tools and equipment to bring with me.* [s20; age 33 years; child 3 months]

> *The app made me less worried about breastfeeding in the public. Before I use the app, I always needed to hide in my car. I ended up going to the same mall again and again. It was troublesome to me and to my family members. Now that I use the app, I have more options. My house is really close to the Emporium shopping mall, but I have never known that there is a feeding room there. I don't have to go to the same department store again and again.* [s10; age 31 years; child 1 year]

> *It brings mothers to a place that everyone is so supportive of breastfeeding. Being there, I didn't have to care what people would think about breastfeeding a 4-year-old child.* [s4; age 39 years; child 4 years]

Intention to Use

After all, although there exist some design issues, these problems did not cause severe inconveniences in using the app. Both quantitative and qualitative results reflected that participants actually found the app acceptable and had a high intention to continue using the app after the trial period. Moreover, they also believed that the app would be beneficial to other mothers in general, as illustrated in the following quotes:

> *My life is easier after using the app, so I will definitely continue using this app.* [s20; age 33; child 3 months]
>
> *I will introduce this app to my friends. It really helps mothers recording data to show doctors.* [s7; age 34; child 10 months]

Discussion

Breastfeeding women from Bangkok and the metropolitan region were recruited to qualitatively examine the usability and usefulness of MoomMae, a mobile phone app designed to promote breastfeeding in Thailand. To maximize the real-life experience, the app was installed on participants' devices and was given to them to use in their daily life for 4 weeks. Individual structured interviews were conducted at the beginning and the end of the trial period. The quantitative results showed a high usability and usefulness score. The qualitative findings provided the insights into usability issues and usefulness of the app.

MoomMae's Role in Supporting Breastfeeding

The Ten Steps to Successful Breastfeeding were developed by WHO and Unicef as a guideline for health facilities to follow to support mothers in breastfeeding [53]. Four of the ten steps (#3, #5, #8, and #10) are not specific to breastfeeding in hospitals and should be continued after discharge. Our findings suggested that MoomMae strongly supports step 8 (encourage breastfeeding on demand) and partly supports step 10 (foster the establishment of breastfeeding support groups and refer mothers to them on discharge).

Self-Efficiency

Our analysis indicates that MoomMae addresses challenges mothers face while breastfeeding. Being a mother is difficult and stressful because of the numerous tasks needed to be achieved daily. MoomMae promotes self-efficiency by helping them to keep things in shape. The numerical values derived from feeding and pumping records help mother to learn their babies and helps manage their schedule. The ubiquitous nature of mobile phones also makes the app more efficient than the notebook approach, as illustrated in the following quote:

> *The app was more convenient. I used to record it on notebooks, and sometimes I was just too lazy to go get my notebook. Using the app is different since I always use my phone while pumping anyways. I can just make a record after I finish pumping.* [s17; age 32 years; child 3 months]

Locality and Public Awareness

The locality of MoomMae is the key to promote breastfeeding in Thailand. MoomMae is the first mobile app that index public breastfeeding places in the local context. We learned from the preuse interviews that it is the most desirable function, and mothers expected it to be the most useful feature. Although the postuse interviews showed that only 11 (out of 21) participants used the feature, participants did not complain about the way the app presented information. Rather, the feedback was regarding the lack of feeding rooms. Moreover, the Thai language used in the app is another key factor that makes MoomMae localized and unique to Thai users, as illustrated in the following quote:

> *The app is really easy to use. I don't find any problems. And, everything is written in Thai, so it's much easier. It's unlike those foreign apps that are written in English.* [s16; age 31 years; child 1 year 6 months]

More importantly, our analysis suggested that MoomMae has the potential to raise public awareness toward breastfeeding. Participants reported that the existence of the app showed them that breastfeeding is so widespread that there is an app developed to support it. One participant described that as more mothers use the app to find feeding rooms, public places would give extra attention to building breastfeeding rooms, and consequently, the public will gradually view breastfeeding as a norm.

Future Implementation

Postuse interviews and The Ten Steps analysis exposed shortcomings of MoomMae that could be improved in future versions. Participants raised the ideas of making the app send a notification to remind them about their pumping schedules, which supports step 5 in The Ten Steps: show mothers how to breastfeed and how to maintain lactation even if they should be separated from their infants. Subjects also want the app to provide question and answer chat rooms, which is associated with step 10: foster the establishment of breastfeeding support groups and refer mothers to them on discharge from the hospital or clinic. Another suggestion from participants is to include information about how to increase or decrease their milk supply, which matches step 5: inform all pregnant women about the benefits and management of breastfeeding. For UI modifications, participants suggested adding cartoon characters and using distinct colors to contrast different functions. From the authors' points of view, our suggestion would be enhancing the existing strength of the app. For instance, as participants found calculating the sum of their feeding records on the app easier, the app can calculate these numbers and display it on the app automatically.

Limitations

We acknowledge that it is difficult to generalize our findings because of small sample size and nonprobability sampling methods. However, prior usability studies have suggested that a small number of participants, as few as 5 to 20, with similar background show great potential for exploring a variety of perspectives that can identify a vast amount of design problems [54,55]. In addition, because of scheduling and transportation constraints of breastfeeding women, it is almost impossible to recruit probability samples. The convenience sampling is, therefore, more feasible and used in previous literature [56-58].

Our study is also limited by the participant characteristics, which may not reflect the whole population of Thai mothers. The exclusive breastfeeding rates of our samples are close to 100%, which is much higher than the rate (12%) of the majority of Thai mothers. We initially targeted to recruit mothers with varied breastfeeding intentions; however, we only received registrations from mothers with high intention in breastfeeding.

This can be explained that mothers with already high intention to breastfeed responded to our recruit posters more quickly and were very interested and responsive in breastfeeding-related studies. Furthermore, our participants were slightly older than the average Thai mothers. In 2015, the mean age of mothers of newborns was 27.5 years [59]; however, our participants gave birth to their current children at the average age of 31.6 years.

We understand that the interview method may not be the ideal protocol for usability evaluation, but it was the most appropriate method after deliberate considerations. For focus group method, the research team thinks that breastfeeding mothers are always busy and have varied schedule to be involved in focus groups. The performance measurement method may not achievable as breastfeeding mothers may need to take care of their children while meeting the researchers, that is, some mothers bring along their children to the research sessions. We also considered that performing a thinking-aloud test in the field may be too noisy and distracting. Taken together, the structured interview was chosen to be the most appropriate method.

Conclusions

Breast milk is universally known to be optimal food for infants. Despite this promise, persisting with breastfeeding is challenging, and exclusive breastfeeding rates remain low in Thailand. To the best of our knowledge, the existing research and commercial apps are country-specific, and there exists a major knowledge gap on how mHealth apps could support breastfeeding in Thailand. MoomMae has therefore been developed and evaluated to support mothers in breastfeeding in the public and in keeping their feeding records. We recruited 21 breastfeeding mothers to participate in preuse and postuse interviews to qualitatively evaluate usability and usefulness of the app.

This study contributes to the growing literature demonstrating how usability assessment of mHealth apps provides invaluable information for iterative developments. The results suggested that MoomMae is easy to use and has the potential to be a useful app for breastfeeding mothers in Thailand. One of the strengths of the app is to support breastfeeding on demand. However, making the flow and inputs more intuitive could enhance the user engagement. Our study also suggests several recommendations for future implementation. Future studies could attempt to recruit participants from a wider range of backgrounds, and other in-depth evaluation methods could be carried out.

Acknowledgments

The authors express their gratitude to the subjects for their participation in this study.

Conflicts of Interest

PC is the principal investigator of MoomMae project.

Multimedia Appendix 1

Screenshots of the “feeding record” function (in Thai).

[PNG File, 927KB - mhealth_v6i1e27_app1.png]

Multimedia Appendix 2

Screenshots of the “breast pump” function (in Thai).

[PNG File, 142KB - mhealth_v6i1e27_app2.png]

Multimedia Appendix 3

Screenshots of the “feeding room” function (in Thai).

[PNG File, 886KB - mhealth_v6i1e27_app3.png]

References

1. Cushing AH, Samet JM, Lambert WE, Skipper BJ, Hunt WC, Young SA, et al. Breastfeeding reduces risk of respiratory illness in infants. Am J Epidemiol 1998 May 1;147(9):863-870. [doi: 10.1093/oxfordjournals.aje.a009540] [Medline: 9583717]
2. Barness LA, Mauer AM, Holliday MA, Anderson AS, Dallman PR, Forbes GB, et al. Commentary on breast-feeding and infant formulas, including proposed standards for formulas. Pediatrics 1976 Feb;57(2):278-285. [Medline: 1250665]
3. Lawrence RA. Breastfeeding: benefits, risks and alternatives. Curr Opin Obstet Gynecol 2000 Dec;12(6):519-524. [Medline: 11128416]
4. Kramer MS, Kakuma R. The optimal duration of exclusive breastfeeding: a systematic review. Adv Exp Med Biol 2004;554:63-77. [Medline: 15384567]

5. UNICEF. New York: United Nations Children's Fund (UNICEF); 2016. The State of the World's Children 2016: A Fair Chance for Every Child URL: https://www.unicef.org/publications/files/UNICEF_SOWC_2016.pdf [WebCite Cache ID 6rpR2Pg6]
6. Thomson G, Ebisch-Burton K, Flacking R. Shame if you do--shame if you don't: women's experiences of infant feeding. Matern Child Nutr 2015 Jan;11(1):33-46. [doi: 10.1111/mcn.12148] [Medline: 25138617]
7. Stuebe A, Bonuck K. What predicts intent to breastfeed exclusively? Breastfeeding knowledge, attitudes, and beliefs in a diverse urban population. Breastfeed Med 2011 Dec;6(6):413-420 [FREE Full text] [doi: 10.1089/bfm.2010.0088] [Medline: 21342016]
8. Grant A. "I…don't want to see you flashing your bits around": Exhibitionism, othering and good motherhood in perceptions of public breastfeeding. Geoforum 2016 May;71:52-61. [doi: 10.1016/j.geoforum.2016.03.004]
9. Vaughn LM, Ireton C, Geraghty SR, Diers T, Niño V, Falciglia GA, et al. Sociocultural influences on the determinants of breast-feeding by Latina mothers in the Cincinnati area. Fam Community Health 2010;33(4):318-328. [doi: 10.1097/FCH.0b013e3181f3b2be] [Medline: 20736758]
10. World Health Organization. Geneva: World Health Organization; 1998. Evidence for the Ten Steps to Successful Breastfeeding URL: http://apps.who.int/iris/bitstream/10665/43633/1/9241591544_eng.pdf [WebCite Cache ID 6rpR6Z2rL]
11. Berridge K, McFadden K, Abayomi J, Topping J. Views of breastfeeding difficulties among drop-in-clinic attendees. Matern Child Nutr 2005 Oct;1(4):250-262. [doi: 10.1111/j.1740-8709.2005.00014.x] [Medline: 16881907]
12. Wood NK, Sanders EA, Lewis FM, Woods NF, Blackburn ST. Pilot test of a home-based program to prevent perceived insufficient milk. Women Birth 2017 Dec;30(6):472-480. [doi: 10.1016/j.wombi.2017.04.006] [Medline: 28529088]
13. Pate B. A systematic review of the effectiveness of breastfeeding intervention delivery methods. J Obstet Gynecol Neonatal Nurs 2009;38(6):642-653. [doi: 10.1111/j.1552-6909.2009.01068.x] [Medline: 19930278]
14. Bonuck K, Stuebe A, Barnett J, Labbok MH, Fletcher J, Bernstein PS. Effect of primary care intervention on breastfeeding duration and intensity. Am J Public Health 2014 Feb;104 Suppl 1:S119-S127. [doi: 10.2105/AJPH.2013.301360] [Medline: 24354834]
15. Jiang H, Li M, Wen L, Hu Q, Yang D, He G, et al. Effect of short message service on infant feeding practice: findings from a community-based study in Shanghai, China. JAMA Pediatr 2014 May;168(5):471-478. [doi: 10.1001/jamapediatrics.2014.58] [Medline: 24639004]
16. Labarère J, Gelbert-Baudino N, Laborde L, Arragain D, Schelstraete C, François P. CD-ROM-based program for breastfeeding mothers. Matern Child Nutr 2011 Jul;7(3):263-272. [doi: 10.1111/j.1740-8709.2009.00235.x] [Medline: 21689269]
17. Fahami F, Mohamadirizi S, Bahadoranffect P. Effect of electronic education on the awareness of women about post partum breast feeding. Int J Pediatr 2014;2:57-63. [doi: 10.22038/ijp.2014.2985]
18. Joshi A, Wilhelm S, Aguirre T, Trout K, Amadi C. An interactive, bilingual touch screen program to promote breastfeeding among Hispanic rural women: usability study. JMIR Res Protoc 2013 Nov 7;2(2):e47 [FREE Full text] [doi: 10.2196/resprot.2872] [Medline: 24200498]
19. Edwards RA, Bickmore T, Jenkins L, Foley M, Manjourides J. Use of an interactive computer agent to support breastfeeding. Matern Child Health J 2013 Dec;17(10):1961-1968. [doi: 10.1007/s10995-013-1222-0] [Medline: 23329167]
20. Salonen AH, Kaunonen M, Astedt-Kurki P, Järvenpää AL, Tarkka MT. Development of an internet-based intervention for parents of infants. J Adv Nurs 2008;64(1):60-72. [doi: 10.1111/j.1365-2648.2008.04759.x]
21. Huang MZ, Kuo SC, Avery MD, Chen W, Lin KC, Gau ML. Evaluating effects of a prenatal web-based breastfeeding education programme in Taiwan. J Clin Nurs 2007 Aug;16(8):1571-1579. [doi: 10.1111/j.1365-2702.2006.01843.x] [Medline: 17655546]
22. Clark A, Anderson J, Adams E, Baker S, Barrett K. Assessing an infant feeding web site as a nutrition education tool for child care providers. J Nutr Educ Behav 2009;41(1):41-46. [doi: 10.1016/j.jneb.2007.12.007]
23. Hannula LS, Kaunonen ME, Puukka PJ. A study to promote breast feeding in the Helsinki Metropolitan area in Finland. Midwifery 2014 Jun;30(6):696-704. [doi: 10.1016/j.midw.2013.10.005]
24. Scott BL. Aquila.usm.edu. The Effectiveness of Internet and Instant Messaging Approaches in Promoting Intention to Breastfeed URL: http://aquila.usm.edu/dissertations/27/ [WebCite Cache ID 6rpRxFBU9]
25. García-Gómez JM, de la Torre-Díez I, Vicente J, Robles M, López-Coronado M, Rodrigues JJ. Analysis of mobile health applications for a broad spectrum of consumers: a user experience approach. Health Informatics J 2014 Mar;20(1):74-84. [doi: 10.1177/1460458213479598] [Medline: 24550566]
26. Play.google. Moms Pump Here URL: https://play.google.com/store/apps/details?id=com.MomsPumpHere.MomsPumpHere&hl=en [accessed 2017-06-18] [WebCite Cache ID 6rpSBQTPD]
27. Play.google. latchME - breastfeed easier URL: https://play.google.com/store/apps/details?id=com.latchme.app [accessed 2017-06-18] [WebCite Cache ID 6rpSNHDr4]
28. itunes.apple. Mamava pump & nursing finder URL: https://itunes.apple.com/us/app/mamava-lactation-suite-locator/id901989849?mt=8 [accessed 2017-06-18] [WebCite Cache ID 6rpSYyNPb]
29. Play.google. Feed Finder URL: https://play.google.com/store/apps/details?id=com.culturelab.feedfinder [accessed 2017-07-09] [WebCite Cache ID 6rpSuGmql]

30. Play.google. BabyPlaces URL: https://play.google.com/store/apps/details?id=de.meltingelements.babyplaces [WebCite Cache ID 6rpSpc939]
31. Play.google. Mamamap URL: https://play.google.com/store/apps/details?id=ch.catatec.mamamap&hl=en [accessed 2017-06-20] [WebCite Cache ID 6rpSvfXbb]
32. Play.google. Nursing Room SG URL: https://play.google.com/store/apps/details?id=com.NursingRoomSG.NursingRoomSG&hl=en [accessed 2017-06-18] [WebCite Cache ID 6rpTcf4Rk]
33. Breastfeedingthai. These places provide “breast milk corner” to facilitate all your mother URL: http://www.breastfeedingthai.com/index.php?lay=show&ac=article&Id=297087 [accessed 2017-06-18] [WebCite Cache ID 6rpTrFiXd]
34. Pandatraveldiary. 2015 Apr 23. Little Panda's Travel Diary URL: https://pandatraveldiary.com/m%E0%B8%BAom-baby-breastfeeding-nursing-room-review/ [accessed 2017-06-17] [WebCite Cache ID 6rpTsNq9q]
35. Khajochi. Introducing “Milk Room” at Central World, big and clean URL: http://www.khajochi.com/2015/10/breast-feeding-room-central-world.html [accessed 2017-07-09] [WebCite Cache ID 6rpTwBI0t]
36. Play.google. Breastfeeding tracker, pump log and baby diary URL: https://play.google.com/store/apps/details?id=com.whisperarts.kids.breastfeeding [accessed 2017-07-09] [WebCite Cache ID 6rpTyKsA1]
37. Play.google. Baby Manager - Breastfeeding Tracker & Community URL: https://play.google.com/store/apps/details?id=innmov.babymanager [accessed 2017-07-09] [WebCite Cache ID 6rpU2sADq]
38. Play.google. Breastfeeding Tracker Baby Log URL: https://play.google.com/store/apps/details?id=com.cg.android.babynursingrel [accessed 2017-07-09] [WebCite Cache ID 6rpU5s9F7]
39. Play.google. Feed Baby - Baby Tracker URL: https://play.google.com/store/apps/details?id=au.com.penguinapps.android.babyfeeding.client.android [accessed 2017-07-09] [WebCite Cache ID 6rpU8QDa5]
40. Play.google. Mamma Baby - Breastfeeding Log URL: https://play.google.com/store/apps/details?id=com.lifenstats.cyrano.mammababy [accessed 2017-07-09] [WebCite Cache ID 6rpUHa4dB]
41. Play.google. Glow Baby Breastfeeding Tracker, Nursing Timer App URL: https://play.google.com/store/apps/details?id=com.glow.android.baby [accessed 2017-06-17] [WebCite Cache ID 6rpUMH8ty]
42. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
43. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
44. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014 Aug 13;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
45. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
46. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May 1;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
47. Daniels J, Fels S, Kushniruk A, Lim J, Ansermino JM. A framework for evaluating usability of clinical monitoring technology. J Clin Monit Comput 2007 Oct;21(5):323-330. [doi: 10.1007/s10877-007-9091-y] [Medline: 17701381]
48. Nielsen J, Molich R. Heuristic Evaluation of User Interfaces. New York, NY, USA: ACM; 1990 Presented at: CHI '90 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 1-5, 1990; Seattle, Washington, USA p. 249-256. [doi: 10.1145/97243.97281]
49. Bertini E, Gabrielli S, Kimani S, Catarci T, Santucci G. Appropriating and assessing heuristics for mobile computing. New York, NY, USA: ACM Press; 2006 Presented at: AVI '06 Proceedings of the working conference on Advanced visual interface; May 23-26, 2006; Venezia, Italy p. 119-126. [doi: 10.1145/1133265.1133291]
50. Watkins I, Kules B, Yuan X, Xie B. Heuristic evaluation of healthy eating apps for older adults. J Consum Health Internet 2014;18(2):105-127. [doi: 10.1080/15398285.2014.902267]
51. Yuan MJ, Finley GM, Long J, Mills C, Johnson RK. Evaluation of user interface and workflow design of a bedside nursing clinical decision support system. Interact J Med Res 2013 Jan 31;2(1):e4 [FREE Full text] [doi: 10.2196/ijmr.2402] [Medline: 23612350]
52. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/BF03195514]
53. BinDhim NF, McGeechan K, Trevena L. Who uses smoking cessation apps? A feasibility study across three countries via smartphones. JMIR Mhealth Uhealth 2014 Feb 6;2(1):e4 [FREE Full text] [doi: 10.2196/mhealth.2841] [Medline: 25098439]
54. Mann D, Riddell L, Lim K, Byrne LK, Nowson C, Rigo M, et al. Mobile phone app aimed at improving iron intake and bioavailability in premenopausal women: a qualitative evaluation. JMIR Mhealth Uhealth 2015 Sep 28;3(3):e92 [FREE Full text] [doi: 10.2196/mhealth.4300] [Medline: 26416479]
55. Fanning J, Mackenzie M, Roberts S, Crato I, Ehlers D, McAuley E. Physical activity, mind wandering, affect, and sleep: an ecological momentary assessment. JMIR Mhealth Uhealth 2016 Aug 31;4(3):e104 [FREE Full text] [doi: 10.2196/mhealth.5855] [Medline: 27580673]

56. Knight BA, McIntyre HD, Hickman IJ, Noud M. Qualitative assessment of user experiences of a novel smart phone application designed to support flexible intensive insulin therapy in type 1 diabetes. BMC Med Inform Decis Mak 2016 Dec 15;16:119 [FREE Full text] [doi: 10.1186/s12911-016-0356-6] [Medline: 27629774]
57. Kushniruk AW, Patel VL, Cimino JJ. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. Proc AMIA Annu Fall Symp 1997:218-222 [FREE Full text] [Medline: 9357620]
58. World Health Organization and Joint United Nations Programme on HIV/AIDS. Protecting, promoting and supporting breastfeeding: the special role of maternity services. Int J Gynaecol Obstet 1990;31 Suppl 1:171-183. [doi: 10.1016/0020-7292(90)90104-S] [Medline: 1972080]
59. Social.nesdb.go.th. 2017. Number and fertility rates Per 1,000 female births by age group, 2002-2016 URL: http://social.nesdb.go.th/SocialStat/StatReport_Final.aspx?reportid=215&template=1R2C&yeartype=M&subcatid=15 [accessed 2017-11-07] [WebCite Cache ID 6u5fEA6jM]

Abbreviations

mHealth: mobile health
NSTDA: National Science and Technology Development Agency
SD: standard deviation
UI: user interface
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.07.17; peer-reviewed by L Dang, T Aguirre, C Adewole; comments to author 31.08.17; revised version received 10.10.17; accepted 23.11.17; published 26.01.18.

Please cite as:
Wang CJ, Chaovalit P, Pongnumkul S
A Breastfeed-Promoting Mobile App Intervention: Usability and Usefulness Study
JMIR Mhealth Uhealth 2018;6(1):e27
URL: http://mhealth.jmir.org/2018/1/e27/
doi:10.2196/mhealth.8337
PMID:29374000

Original Paper

Tanzania Health Information Technology (T-HIT) System: Pilot Test of a Tablet-Based System to Improve Prevention of Mother-to-Child Transmission of HIV

Sheana Bull[1], MPH, PhD; Deborah SK Thomas[2], MA, PhD; Elias C Nyanza[3], MPH; Sospatro E Ngallaba[3], MSc, MPH, MD

[1]Department of Community and Behavioral Health, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

[2]Department of Geography & Environmental Sciences, University of Colorado Denver, Denver, CO, United States

[3]School of Public Health, Catholic University of Health and Allied Sciences, Mwanza, United Republic Of Tanzania

Corresponding Author:
Deborah SK Thomas, MA, PhD
Department of Geography & Environmental Sciences
University of Colorado Denver
PO Box 173364, CB 172
Denver, CO, 80217-3364
United States
Phone: 1 303 315 7557
Email: deborah.thomas@ucdenver.edu

Abstract

Background: The prevention of mother-to-child transmission (PMTCT) of HIV requires innovative solutions. Although routine monitoring is effective in some areas, standardized and easy-to-scale solutions to identify and monitor pregnant women, test them for HIV, and treat them and their children is still lacking. Mobile health (mHealth) offers opportunities for surveillance and reporting in rural areas of low- and middle-income countries.

Objective: The aim of this study was to document the preliminary impacts of the Tanzania Health Information Technology (T-HIT) system mHealth intervention aimed at health workers for PMTCT care delivery and capacity building in a rural area of Tanzania.

Methods: We developed T-HIT as a tablet-based system for an electronic data collection system designed to capture and report PMTCT data during antenatal, delivery, and postnatal visits in Misungwi, Tanzania. T-HIT was tested by health workers in a pilot randomized trial comparing seven sites using T-HIT assigned at random to seven control sites; all sites maintained standard paper record-keeping during the pilot intervention period. We compared numbers of antenatal visits, number of HIV tests administered, and women testing positive across all sites.

Results: Health workers recorded data from antenatal visits for 1530 women; of these, 695 (45.42%) were tested for HIV and 3.59% (55/1530) tested positive. Health workers were unable to conduct an HIV test for 103 women (6.73%, 103/1530) because of lack of reagent, which is not captured on paper logs. There was no difference in the activity level for testing when comparing sites T-HIT to non-T-HIT sites. We observed a significant postintervention increase in the numbers of women testing positive for HIV compared with the preintervention period (P=.04), but this was likely not attributable to the T-HIT system.

Conclusions: T-HIT had a high degree of acceptability and feasibility and is perceived as useful by health workers, who documented more antenatal visits during the pilot intervention compared with a traditional system of paper logs, suggesting potential for improvements in antenatal care for women at risk for HIV.

(JMIR Mhealth Uhealth 2018;6(1):e16) doi:10.2196/mhealth.8513

KEYWORDS

mHealth; decision aids; HIV; healthcare workers

Introduction

Rationale

According to the World Health Organization (WHO), an estimated 36.7 million people globally live with human HIV, and nearly 70% of them reside in sub-Saharan Africa [1]. Although there has been substantial progress in the fight against HIV, we fall short of global goals to eliminate new infections among children and substantially reduce acquired immune deficiency syndrome–related maternal deaths [2]. Tanzania, with an estimated 5.6% HIV prevalence among pregnant women attending antenatal care (ANC) clinics [3], was among 22 countries targeted by the United Nations Global Plan to eliminate new HIV infections in children [2]. The Global Plan sets a target for reducing the mother-to-child transmission of HIV during pregnancy and breastfeeding to less than 5%, although transmission in many countries remains stubbornly high [2]. Key elements of prevention of mother-to-child transmission (PMTCT) include: testing pregnant women for HIV, initiating lifelong antiretroviral (ARV) therapy for women with HIV (called Option B+, promoted via WHO guidelines in 2010) [4], delivering in a care facility, initiating nevirapine therapy for exposed infants, diagnosing infants, and offering ongoing linkages to care.

In Tanzania, the primary challenges include limited ANC clinic attendance across the four recommended WHO visits and limited integration between PMTCT and broader maternal and child health services [5,6]. In sharp contrast to the WHO goals, only an estimated 76% of HIV positive pregnant women attending ANC clinic also received PMTCT services, and only 43% of HIV exposed infants received ARV drugs [5,6]. Although routine monitoring is effective in some areas, we continue to lack a standardized and easy-to-scale solution to identify and monitor pregnant women, test them for HIV, and treat them and their children.

Mobile platforms such as phones and tablets have tremendous potential to impact health care delivery and health outcomes. The use of cell phones worldwide has expanded rapidly over the past decade in both developed and developing countries. By the end of 2015, there were nearly 7.1 billion mobile cellular subscriptions globally, and close to 100% of the population was covered by a mobile signal, a drastic increase from 20% coverage in 2003 [7], even reaching poor-resource settings [8]. The universality of cell phones provides an opportunity for their use in broad and scale-up of technology-based health interventions, particularly in developing and resource-poor areas. A proliferation of innovations that integrate the use of mobile and wireless devices to improve health outcomes, health care services, and health research into care delivery, often called *mHealth*, has occurred concomitantly with the growth of cell phone usage [9].

Researchers have implemented mobile health (mHealth) apps in a range of settings and multitude of health targets [10] for facilitation of care delivery, medical records charting, patient and health worker education, disease prevention, and patient self-management. These tools can improve surveillance, clinical care, prevention, and self-management. Furthermore, they have the potential to expand population-level public health impacts through wider dissemination and scale-up for wide spread use [11]. Successful mHealth interventions can often intensify their effects when they are guided by behavioral and social science theory to help in the design, implementation, and analysis of effects [12].

Although mHealth has more often focused on prevention and self-management for behavior change at the individual level, attention has broadened toward targeting the health care worker in rural and resource-poor settings as a possible sustainable intervention model. Our own recent report offered a synthesis of findings from 31 peer-reviewed studies related to the use of mHealth solutions for delivery of health care in resource-limited settings [13]. Overall, the findings demonstrate a substantial benefit to health workers, their patients, and care delivery systems when mobile technology tools such as mobile phones and tablets are used. Acceptability of these tools for care delivery is high, and evidence shows that the use of mHealth tools can improve communication between health workers and their patients, health workers and clinic staff, as well as between health workers and their supervisors. Use of mHealth tools by health workers is associated with improved compliance with treatment protocols among patients and improved health outcomes. mHealth tools are used successfully in surveillance efforts to improve quality and efficiency of data collection [13].

Objective

The Tanzania Health Information Technology (T-HIT) system was developed as an easy-to-use tablet-based patient data collection mHealth solution that could potentially be scaled to health facilities throughout the region with the goal of improving the continuity and quality of care for PMTCT delivery to women in a rural, resource-limited setting. In addition to improved and integrated patient records that can enhance continuity of care through the linking of patient records over time, electronic record-keeping also enables the documentation of trends and patterns in overall care delivery. Thus, T-HIT has the potential to enhance care delivery and function as a decision-support tool for health workers and administrators. This paper aims to (1) document T-HIT system use and describe health worker clinical activities at the health facility level for PMTCT, (2) demonstrate T-HIT feasibility for use by health workers over a 3-month period, and (3) document system capacity for HIV testing and identification of new HIV infections.

Methods

Study Design

This was a randomized controlled pilot study. Working in Misungwi District in Northwestern Tanzania, we identified fourteen health facility sites within the district that represented a mix of hospital, health centers, and dispensaries from the total of 42 health facilities. The Tanzanian health system, which is structured similar to many other sub-Saharan countries, is organized in a pyramid referral structure, with dispensaries being the smallest, then health centers, district hospitals, and finally regional hospitals. All health facilities in the district were eligible to participate.

To ensure equivalent resource distribution across intervention and control sites, we matched pairs and planned for the intervention and control groups to each have a mix of hospital (one intervention or one control), health centers (two intervention or two control), and dispensaries (four intervention or four control). This includes all hospitals and all health centers in the district and eight of the remaining 36 dispensaries as an intervention or control site in the study. Stratification included (1) Facilities that were closer (less than 20 kilometers) or further (greater than 20 kilometers) and (2) Dispensaries with more prior experience (having PMTCT during 2010 preliminary work) or less prior experience (having no PMTCT during 2010 preliminary work). Once matched, we randomly assigned a facility within each pair to either intervention (training for, and use of, the T-HIT system) or control (standard of care, which is recording visits in handwritten paper logs).

Development of the Tanzania Health Information Technology System

The focus of the T-HIT system was on strengthening PMTCT through an mHealth tablet-based solution for health workers that acts as a platform for capturing patient data and providing some immediate feedback to health workers to enhance care. To that end, T-HIT was designed with three primary features to support the electronic capture of data relevant for PMTCT outcomes. First, it facilitates data collection during a clinic visit for PMTCT in a way that mimics the data elements captured in paper records. Second, it also serves to help reinforce health worker clinical decisions with integrated alerts and reminders about specific protocols. Finally, a reporting dashboard relays composite data for each health facility and for the district (in this instance all seven facilities) in near real time. With paper records, reporting is much more limited. Importantly, T-HIT interface was designed with feedback from health workers and the district health management team.

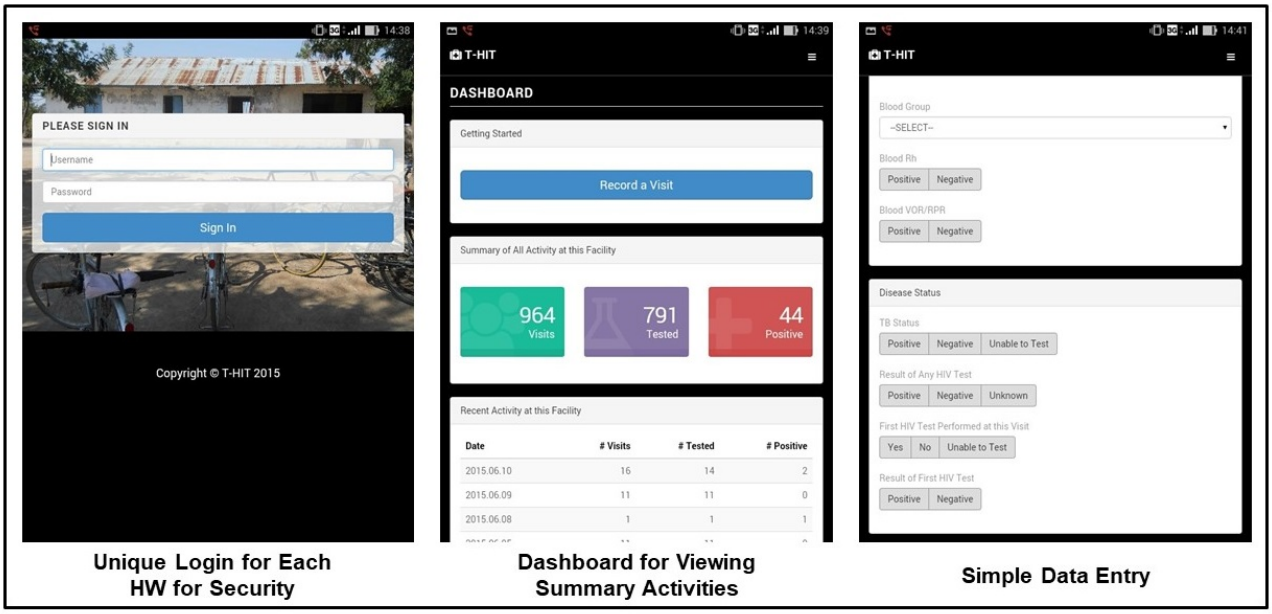
For this initial pilot, the T-HIT system was designed to document delivery of a subset of elements traditionally documented within an antenatal face-to-face visit for a mother with, or at risk of HIV, and would typically be recorded in paper logs maintained at the health facility and on the maternal and child health (MCH) card that a woman keeps with her. Because the goal was, in part, to ensure compatibility with current work flows, nearly no changes in data elements were made to those collected and charted within the system compared with the paper standard of care. The T-HIT interface was designed to capture these data elements with ease-of-use in mind, using clean and uncluttered interfaces with drop-down menus and check box options to limit errors in data entry (see Figure 1).

Data Collection

An inventory of all facilities in the district was conducted in summer 2014 and updated in January 2015 to establish cell phone connectivity and availability of electricity, as well as to develop base maps of facilities and roads for use in descriptive mapping of the data across the district. All facilities in the district had cell phone connectivity, though with variation in the network operator available. For those facilities without electricity, a solar charger was provided with the tablet. In addition to the base maps, the primary analysis presented in this paper uses two data sources: (1) paper logs for both T-HIT and control sites and (2) T-HIT entries.

Paper registries are the standard of care at health facilities across Tanzania and much of sub-Saharan Africa; health workers handwrite data for all mothers receiving care at the time of a visit in facility logs and on the maternal health card. Study research assistants collected data from all fourteen intervention and control sites relevant to PMTCT outcomes from patient antenatal registries for a 20-week (5-month) period before initiation of T-HIT and for the 12-week (3-month) intervention period when T-HIT was being used. Data were also collected through the T-HIT system for the seven sites from February 23, 2015 to May 23, 2015.

Figure 1. Selected screenshots of the Tanzania Health Information Technology (T-HIT) system.



Health workers at T-HIT sites continued maintaining handwritten records to adhere to the standard of care, as well as documenting care delivery via the T-HIT tablets. A unique identification number was generated for each patient to protect privacy and confidentiality.

Data Analysis

This analysis focuses on documenting HIV testing in the intervention communities, as well as comparison of the total number of antenatal visits, HIV tests, and women testing positive for HIV over time in T-HIT and non-T-HIT sites. Although T-HIT captured antenatal, delivery, and postnatal data, the modeling only utilized the antenatal data as this corresponded to the antenatal paper records collected at intervention and control sites. Furthermore, we utilized the total visits from the T-HIT system, rather than unique patient visits, in the analysis because of an inability to link records in the paper patient logs. Linking patient data in the facility logs would require a manual record-by-record evaluation, looking for the name of each patient in every subsequent record during the pre- and postintervention phases, which was not feasible. Importantly, in the context of documenting trialability, the total number of visits reflects workload.

The analyses included descriptive data summary and mapping along with multivariate modeling. We generated total numbers from the T-HIT system to illustrate the level of productivity. These data were, in turn, mapped using *ArcGIS version 10.2* (Esri, Redlands, CA) geographic information system (GIS) software to offer visual depictions of HIV incidence and testing distribution among the T-HIT intervention sites. Generalized estimating equations (GEEs) in *SAS version 9.4* (SAS Institute Inc, Cary, NC) were used to model the effect of time (week number), phase (preintervention vs postintervention), and record type (T-HIT electronic, T-HIT paper, or non-T-HIT paper). A weekly visits rate ratio, as well as odds ratio was used to determine if there were greater odds of being seen in T-HIT or non-T-HIT sites.

Segmented mixed-model Poisson regression with an autoregressive (AR[1]) structure for repeated measures on facilities over weeks were used, trying several functional forms by week, including linear, quadratic, cubic, and categorical. In no case was the time effect significant, so the simplest functional form, linear, was used in the final model. Additionally, interaction terms were initially included to allow the effect of time to vary by phase, the effect of group to vary by phase, and the effect of time to vary by group and phase (the group×phase×week interaction). These interactions were removed from the model one at a time using backwards elimination. In the end, there were no significant interactions, although the week×phase interaction remained in the model to preserve the segmented nature of the model.

GEEs were also used to model the effect of time (week number), phase (pre intervention vs post intervention), and record type (T-HIT electronic, T-HIT paper, or Non-T-HIT paper) on the odds of (1) being tested for HIV and (2) testing positive for HIV using segmented mixed model logistic regression with an AR(1) structure for repeated measures on facilities over weeks. Again, several functional forms for week were tried, including linear, quadratic, cubic, and categorical. In no case was the time effect significant, so the simplest functional form, linear, was used in the final model. Additionally, interaction terms were initially included to allow the effect of time to vary by phase, the effect of group to vary by phase, and the effect of time to vary by group and phase (the group×phase×week interaction) for each model. These interactions were removed from the model one at a time using backwards elimination. In the end, there were no significant interactions, although the week×phase interaction remained in the model to preserve the segmented nature of the models.

Ethical Statement

Ethical approval was obtained from the Conjoint Catholic University of Health and Allied Sciences and Bugando Medical Centre Research Review and Ethics Committee (Ref. *CREC/051/2013*), the Tanzania National Institute for Medical Research (*NIMR/HQ/R.8a/Vol.IX/1662*), and the Colorado Multiple Institutional Review Board (Protocol 13-2166). Permission to conduct research in Tanzania was also obtained from the Tanzania Commission for Science and Technology and regional and district authorities in Mwanza and Misungwi, respectively. Health worker participation in the study was voluntary, and all who were approached agreed to take part.

Results

Table 1 presents descriptive data on facilities and mean number of visits, HIV tests, and numbers testing positive for HIV from the T-HIT electronic records. Health workers used T-HIT to document 1594 antenatal visits across 1530 unique patients, with 558 of these visits at the district hospital. There were between 69 and 226 antenatal visits to dispensaries and health centers and an average of 16.27 antenatal visits each week across all T-HIT sites. Almost all women (96.01%, 1469/1530) had only one antenatal visit, and only 56 women in the pilot had more than one antenatal visit. We documented outcomes from 695 HIV tests that took place during an antenatal visit in the pilot period; of these, there were 55 women who tested positive (3.59% of the 1530 unique women seen during the pilot). Importantly, T-HIT documented if a health worker was unable to conduct an HIV test, most likely because of a lack of reagent. T-HIT documented 103 women for whom health workers were unable to successfully complete an HIV test antenatally (6.73%, 103/1530 of the unique women seen). Those facilities with five or less visits are not reported for confidentiality purposes, although the data were included in the model.

Paper records at the T-HIT sites recorded significantly fewer antenatal visits during the same period (N=879), though similarly proportionately lower at all facilities except Misasi Health Centre, which was the only site to record a higher number in the paper records. The total number of HIV tests (N=486) and positive HIV results (N=22) were also substantially lower than the T-HIT electronic records. Table 2 presents the paper records for the intervention period for the non-T-HIT control sites. The total number of visits (N=866), HIV tests (N=343), and HIV positive results (N=25) are comparable between the paper control and intervention records.

Table 1. Antenatal visit information at Tanzania Health Information Technology (T-HIT) health facility sites.

Health facility sites	Misungwi District Hospital	Misasi Health Center	Mwawile Dispensary	Mbarika Health Center	Mondo Dispensary	Nguge Dispensary	Gambajiga Dispensary	Total T-HIT
Electronic records								
Total visits	558	69	187	226	205	175	164	1594
Mean per week (SD[a])	40.6 (15.6)	4.9 (5.5)	13.4 (8.8)	16.1 (11.0)	14.6 (8.8)	12.5 (8.1)	11.7 (6.2)	16.27 (3.38)
Total HIV tests	193	6	97	121	39	139	100	695
Mean per week (SD)	14.1 (11.0)	0.46 (1.13)	6.9 (5.2)	9.0 (6.7)	2.5 (3.6)	10.0 (7.8)	7.7 (4.1)	7.2 (3.2)
Total unable to complete an HIV test	23	__[b]	24	9	34	10	0	103
Mean per week (SD)	1.6 (3.5)	—	1.7 (3.6)	0.6 (1.2)	2.4 (2.7)	0.7 (2.2)	0 (0)	1.1 (1.4)
Total testing HIV+	26	—	10	—	0	6	—	55
Mean per week (SD)	1.9 (1.3)	—	0.7 (1.4)	—	0.0 (0.0)	0.4 (0.5)	—	0.6 (0.5)
Paper records								
Total visits	307	258	79	104	34	65	32	879
Mean per week (SD)	29.63 (9.49)	22.63 (7.71)	8.07 (3.73)	9.81 (4.9)	3.00 (1.92)	6.64 (2.65)	3.92 (2.43)	12.50 (11)
Total HIV tests	202	56	64	68	6	58	32	486
Mean per week (SD)	22.50 (10.58)	4.81 (6.75)	6.47 (4.03)	6.81 (5.74)	0.64 (1.34)	6.14 (2.93)	3.85 (2.44)	7.58 (8.75)
Total testing HIV+	8	—	7	—	0	—	—	22
Mean per week (SD)	0.69 (0.95)	—	0.73 (0.88)	—	0.00 (0)	—	—	0.35 (0.69)

[a]SD: standard deviation.

[b]Five or less visits reported; totals were included in the model.

Table 2. Antenatal visit information at comparison health facility sites, intervention phase.

Health facility sites	Bukumbi Mission Hospital	Busongo Health Center	Igokelo Dispensary	Koromije Health Center	Lubiri Dispensary	Nyamijundu Dispensary	Nyang homango Dispensary	Total paper cohort
Paper records								
Total visits	178	137	94)	179	70	131	77	866
Mean per week (SD[a])	10.81 (3.67)	10.60 (3.25)	5.94 (3.51)	16.1 (11.0)	5.57 (3.41)	12.5 (8.1)	6.29 (3.65)	8.91 (5.06)
Total HIV tests	95	34	14	71	27	47	55	343
Mean per week (SD)	7.69 (3.65)	1.93 (3.47)	0.19 (0.4)	4.06 (5.18)	2.00 (2.88)	10.0 (7.8)	8.08 (5.33)	4.20 (4.65)
Total testing HIV+	21	__[b]	—	6	0	—	—	25
Mean per week (SD)	0.56 (0.89)	—	—	0.31 (0.79)	0.00 (0)	—	—	0.24 (0.6)

[a]SD: standard deviation.

[b]Five or less visits reported; totals were included in the model.

Figure 2. Maps demonstrating testing access and HIV incidence at Tanzania Health Information Technology (T-HIT) study sites.

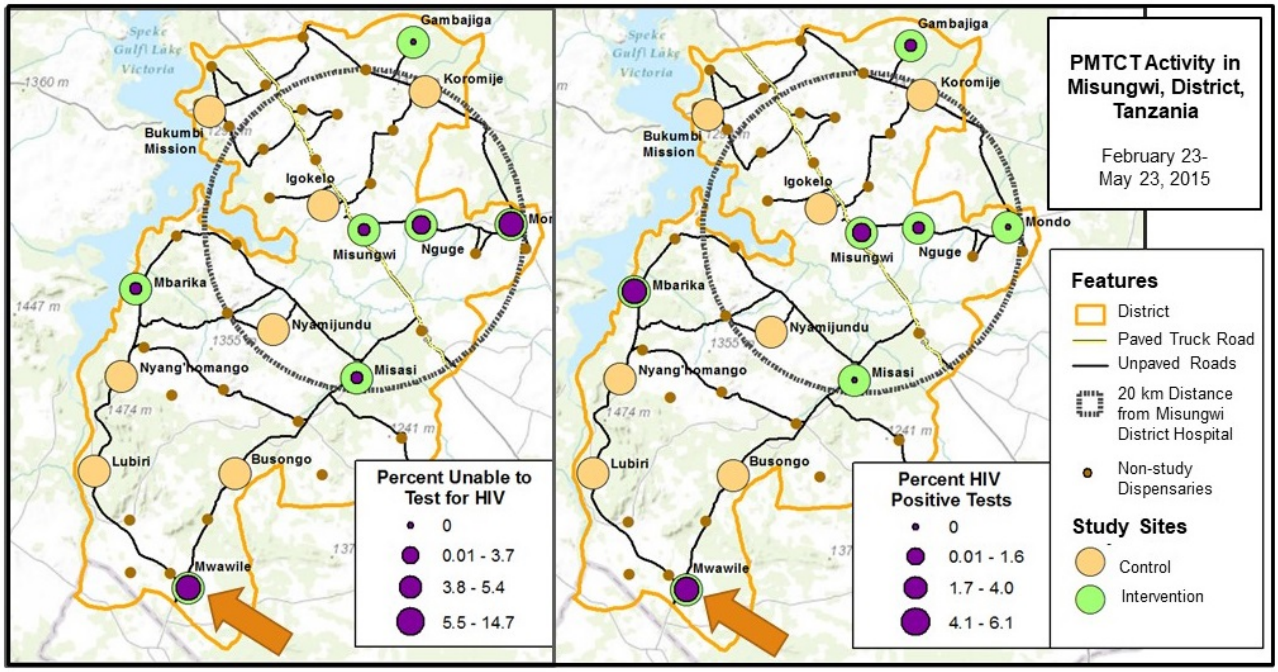


Figure 2 offers a mapped depiction of the distribution of intervention and control sites, along with HIV incidence and testing patterns in intervention sites. The small orange-brown dots represent the location of 28 dispensaries not randomized into the study. The large orange circles are the control sites and the green circles, the intervention sites. The graduated purple circles inside the intervention sites show the percentage of HIV positive tests captured in the T-HIT system on the right map and the percentage of unable to test. Bukumbi Mission and Misungwi are the two hospitals and Misassi, Koromije, Mbarika, and Busongo the health centers. As highlighted with an arrow at the bottom of each map, one site, Mwawile Dispensary, emerged as having both a high number of people unable to test for HIV (shown on the left) and simultaneously a high incidence of HIV among women testing (shown on the right). Further investigation with clinic staff in Mwawile Dispensary revealed that the clinic ran out of reagent for testing during this period, and health workers were unable to complete HIV testing in 12.8% (24/187) of cases.

Table 3 shows rate ratios for total antenatal visits and odds ratios for testing and testing positive for HIV, comparing T-HIT electronic records with paper records at non-T-HIT sites and comparing visits in the preintervention phase to the intervention phase. The number of *weekly antenatal visits* (patients seen) did not differ significantly from week-to-week before the intervention (P=.14). This weekly rate of change did not differ significantly in the postimplementation phase with respect to the preimplementation phase (P=.83), and the implementation itself did not cause any change in the number of patients seen per week (P=.75).

The odds of *being tested* for HIV did not differ significantly from week-to-week before the intervention (Table 2, P=.35). We observed a drop in HIV testing comparing the preintervention with the postintervention period, where women were 0.06 times as likely to be tested after the pilot implementation as before (P=.01).

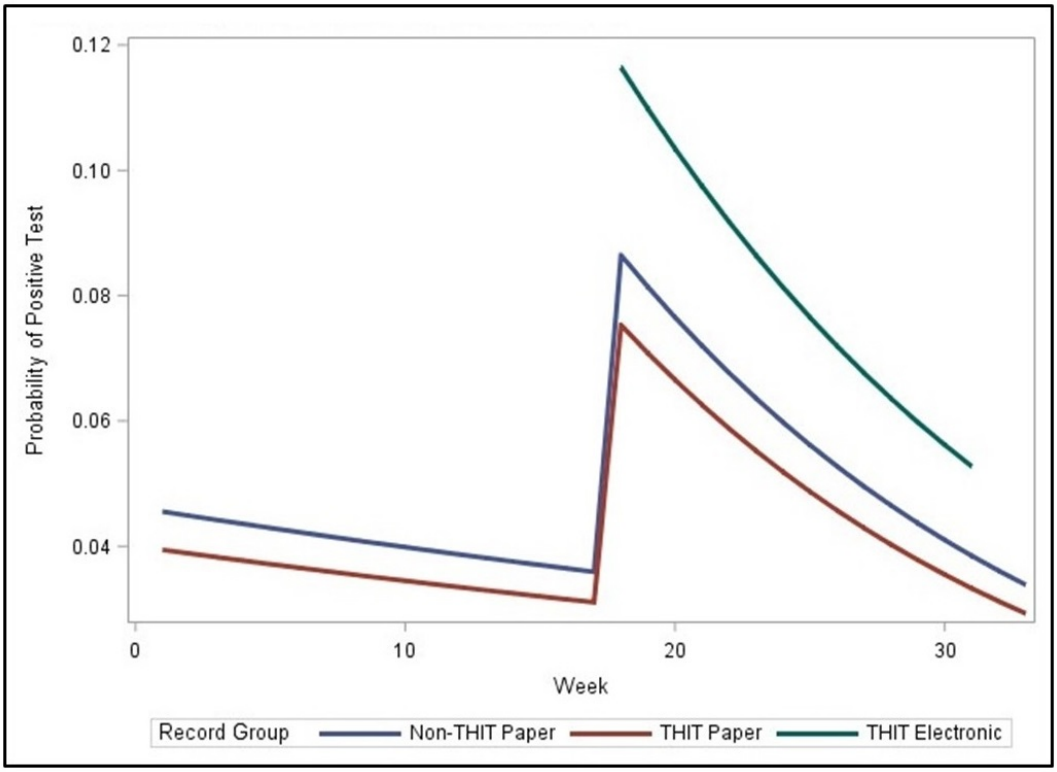
The odds of *testing positive* for HIV did not differ significantly from week-to-week before the intervention (P=.55). Specifically, the odds of testing positive for HIV dropped 2% (95% CI: a drop of 6% to an increase of 4%). However, the weekly rate of change did differ significantly in the postimplementation phase with respect to the preimplementation phase (P=.04). The weekly rate of change in the postimplementation phase was 5% lower than that in the preimplementation phase (95% CI: 9% decrease to 0% increase), giving a significant postintervention weekly rate of change (P=.004), with the odds of testing positive dropping by 6% (95% CI: 11% drop to 2% drop) in the postimplementation phase. This is represented graphically in Figure 3.

Table 3. Model outcomes for rate and odds ratios of patient visits, tested for HIV, and tested positive for HIV; Tanzania Health Information Technology (T-HIT) electronic records compared with paper records in T-HIT sites and comparison sites.

Variable and category	Rate ratio (95% CI)	*P* value
Rate ratios for the number of patients seen weekly		
Week	0.99 (0.97-1.00)	.14
Record group		.20[a]
T-HIT electronic versus non-T-HIT paper (week 18)	1.52 (0.87-2.66)	.14
T-HIT electronic versus T-HIT paper (week 18)	1.14 (0.67-1.95)	.63
T-HIT paper versus non-T-HIT paper (week 18)	1.11 (0.61-2.01)	.74
Phase		
Post intervention versus pre intervention	1.04 (0.71-1.53)	.83
Week×phase interaction		
Post intervention versus pre intervention	1.00 (0.98-1.03)	.75
Odds ratios for the number of patients who were tested for HIV		
Week	0.95 (0.86-1.06)	.35
Record group		.77[a]
T-HIT electronic versus non-T-HIT paper (week 18)	0.89 (0.27-2.90)	.85
T-HIT electronic versus T-HIT paper (week 18)	0.65 (0.24-1.79)	.41
T-HIT paper versus non-T-HIT paper (week 18)	1.28 (0.26-6.16)	.76
Phase		
Post intervention versus pre intervention	0.06 (0.01-0.29)	.001
Week×phase interaction		
Post intervention versus pre intervention	1.10 (0.97-1.23)	.13
Odds ratios for the number of those tested who were positive for HIV		
Week×phase interaction		
Post intervention versus pre intervention	0.95 (0.91-1.00)	.04
Preintervention effect of week	0.98 (0.94-1.04)	.55
Postintervention effect of week	0.94 (0.89-0.98)	.004
Record group		.69[a]
T-HIT electronic versus non-T-HIT paper (week 18)	0.67 (0.24-1.82)	.43
T-HIT electronic versus T-HIT paper (week 18)	0.78 (0.23-2.63)	.68
T-HIT paper versus non-T-HIT paper (week 18)	0.41 (0.19-0.89)	.02
Phase		
Post intervention versus pre intervention	6.43 (2.26-18.35)	.001

[a]Type 3 *P* values.

Figure 3. Odds of testing positive for HIV.



Discussion

Principal Findings

Although there is high acceptability of mHealth solutions in low and middle-income countries (LMIC), evidence for intervention efficacy is limited in LMIC settings [14]. Overall, the evidence from this study suggests that health workers can, and will, use the T-HIT system to record PMTCT visit data. Health workers in the T-HIT hospital, health centers, and dispensaries all used the system without overt challenges. Due to strong indications for relative advantage and feasibility, T-HIT has the potential for scaling. Furthermore, T-HIT may well contribute to increases in PMTCT visits if the quality of care by health workers improves through a system that supports decision making. This alone is a reason to implement and test T-HIT efficacy on a larger scale. We also anticipate dividends beyond the capacity to document in this pilot that can be realized in a larger trial of T-HIT. For example, more immediate information sharing from the T-HIT system to the district hospital could better distribute reagent for HIV testing, shorten time frames to distribute ARV, and better plan for hospital delivery and infant care for HIV positive mothers.

The data suggest that the intervention and control sites were comparable in terms of overall visits and HIV testing for mothers before and during the study period. The T-HIT system captured potentially critical data on the inability to test for HIV, revealing the need for redistribution of HIV reagent much more immediate at some sites. This is a particularly pressing need for clinics with higher HIV incidence, as most women do not return for follow-up antenatal visits when a follow-up test could occur.

The integration of data across visits for mothers is another critical contribution of an electronic data entry mHealth solution. In the paper record system, there are separate logs for antenatal visits, delivery visits, and postnatal visits and yet another record kept by the mother in the form of the MCH card. The T-HIT system has the potential for linking a patient's record across numerous visits. In this way, a woman's health status can be tracked, including risk factors and HIV testing status throughout pregnancy, delivery, and postnatal care. This, in turn, greatly improves opportunities for continuity of care. In an expanded and longer T-HIT trial, monitoring patient-level visits in combination with improved care reinforced by T-HIT decision-support would likely yield an improvement in the number of pre- and postnatal visits, along with an increase in hospital delivery numbers.

A decline in testing for HIV was observed when comparing the intervention period to the preintervention period across all sites, suggesting something other than the newly introduced T-HIT system affected declines in HIV testing. Despite this decline in testing, there was an overall increase in the number of women testing positive for HIV during the pilot study implementation period compared with the preimplementation, with no difference in intervention compared with control sites. This, at least, suggests that whatever the background driver was, the trend was captured in both the paper logs and T-HIT, indicating health workers utilized the interface successfully and these data were appropriately captured by the system, providing evidence for feasibility. In a longer trial, we also posit an improvement for HIV testing capacity and volume resulting from the ability of the health delivery system to effectively respond to real-time stock outages such as HIV reagent.

The odds of testing positive for HIV suggest that women in T-HIT were not more likely to have testing in this short pilot; this suggests an uptick in the numbers of women documented with HIV immediately after the intervention was implemented (regardless of whether it was documented on a paper log or the T-HIT system) that reverted to preintervention levels over time. This may represent a Hawthorne effect, or possibly a temporal factor such as weather or seasonal events. At a minimum, this also provides additional evidence in support of feasibility of the mHealth solution.

T-HIT consistently recorded a higher number of visits, HIV tests, and HIV positive results. We hypothesize that this is quite possibly because the T-HIT system was capturing data that would be found in separate paper log books. Consequently, T-HIT might be already exhibiting the capability for integrating data sources. However, this could also be a result of health workers entering data after the visit, or entering additional data to show an increase of activity knowing that the totals were being displayed in near real time. However, it would be challenging to consistently adjust over 3 months and similarly across facilities. In one instance, Misasi Health Centre entered far fewer records than would be expected entered into T-HIT based both on a comparison to paper records at that site and to the activity at other health centers. Thus, this is likely evidence of a facility that did not fully adopt or implement the T-HIT system with regularity.

We have yet to see widespread adoption of mHealth solutions for care delivery. This suggests incentives are needed to facilitate adoption and use that are targeted at various components of the health care system. For example, incentives can be aimed at the health worker through training or monetary compensation. Perhaps one of the most important incentives is for the system to have perceived usefulness in terms of increasing confidence for providing care. Additionally, policies that obligate use can be established at the systems level. However, before policies that require use of mHealth tools can be realistically established, a careful assessment is needed to ensure organizational readiness to train users and offer technical support for devices and data management. Data quality (completeness and accuracy) issues persist for PMTCT, and although opportunities for improvement are complex, perceived usefulness along with support and supervision are necessary to produce reliable and informative [15]. Long-term sustainability can likely not rely on remuneration for users in low-resource settings and so, multi-faceted approaches for implementation and scaling will be required.

As implementation of mHealth solutions for care delivery increases, a critical consideration of costs associated with technology infrastructure will be required to evaluate whether investment in this infrastructure is warranted. Existing more “low-tech” approaches to data collection, such as pen and paper, may be sufficient. However, if decision makers determine that infrastructural investment in technology for health care delivery is appropriate, then attention to multiple areas to maximize this investment is needed. Many careful considerations are necessary, including technology or equipment choices (computers, servers, phones, and tablets), sufficient staff who can program and maintain such equipment, development of protocols and training programs for health care workers to effectively use technology, development of policies and incentives to motivate use, and attention to regular process evaluations to ensure efficiency and quality in data collection and communication.

Conclusions

This pilot study documented the successful use of the tablet-based mHealth T-HIT system, demonstrated its feasibility for the effective use by health workers, and illustrated the capacity for capturing, facilitating, and monitoring HIV testing and new infections for PMTCT. Furthermore, findings indicate T-HIT offers a potentially scalable mHealth solution for PMTCT data collection to facilitate decision support in resource-scarce settings. Importantly, although the T-HIT system was designed to directly address PMTCT, it could be easily adapted to facilitate monitoring and integrated care delivery for any number of priority health concerns such as antenatal or postnatal care, tuberculosis care, malaria, and chronic conditions.

Acknowledgments

This study was funded by the National Institutes of Health (NIH Project: R21 MH099942-01A1). The analysis was supported in part by NIH/NCRR Colorado CTSI Grant Number UL1 RR025780. Contents are the authors' sole responsibility and do not necessarily represent official NIH views. The authors acknowledge the Catholic University of Health and Allied Sciences in Tanzania, the University of Colorado Denver, and Misungwi District Council in Tanzania in their support of this project. A special thanks goes to the research assistants, Annastazia Magongo, and Sarah Temba for their efforts on this project. The authors gratefully acknowledge the support of administrators, staff, patients, and community members in the Misungwi District of Tanzania where they implemented this project.

Authors' Contributions

SB contributed to the study design, data analysis, data interpretation, and in drafting the manuscript. DSKT contributed to the study design, development of the T-HIT system, mapping activities, data analysis, interpretation of the findings, as well as the writing of the manuscript. ECN contributed in designing the study, developing the survey, supervising the data collection, and reviewing of the manuscript. SEN assisted in study design and reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. World Health Organization. 2016. Number of people (all ages) living with HIV estimates by WHO region URL: http://apps.who.int/gho/data/view.main.22100WHO?lang=en [WebCite Cache ID 6w10OYnW9]
2. Joint United Nations Programme on HIV/AIDS. Global Report: UNAIDS report on the global AIDS epidemic 2013. 2013. URL: http://www.unaids.org/sites/default/files/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf [WebCite Cache ID 6w10bGJKH]
3. Manyahi J, Jullu BS, Abuya MI, Juma J, Ndayongeje J, Kilama B, et al. Prevalence of HIV and syphilis infections among pregnant women attending antenatal clinics in Tanzania, 2011. BMC Public Health 2015 May 22;15(501). [doi: 10.1186/s12889-015-1848-5] [Medline: 25994129]
4. World Health Organization. Avert. 2017. Prevention of Mother-to-Child Transmission (PMTCT) of HIV URL: http://www.avert.org/professionals/hiv-prevention-programming/prevention-mother-child/guidelines [accessed 2017-12-27] [WebCite Cache ID 6w11gCIPO]
5. Tanzania Commission for AIDS (TACAIDS). Country Planning Cycle Database. 2013. Tanzania third national multi-sectoral strategic framework for HIV and AIDS (2013/14-2017/18) URL: http://www.nationalplanningcycles.org/sites/default/files/country_docs/Tanzania/nmsf-iii_eng_final_report_2013mail.pdf [accessed 2017-12-27] [WebCite Cache ID 6w11zo5PC]
6. Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Comission (ZAC), National Bureau of Statistics (NBS), Office of Chief Government Statistician (OCGS), ICF International. Dhsprogram. 2013. HIV/AIDS and Malaria Indicator Survey, 2011-12 URL: https://dhsprogram.com/pubs/pdf/AIS11/AIS11.pdf [accessed 2017-12-27] [WebCite Cache ID 6w12ISy9o]
7. International Telecommunication Union (ITU). Measuring the Information Society Report. Geneva; 2015. URL: https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2015/MISR2015-w5.pdf [WebCite Cache ID 6w12RpFEB]
8. Betjemen TJ, Sogoian SE, Foran MP. mHealth in Sub-Saharan Africa. Int J Telemed Appl 2013;2013. [doi: 10.1155/2013/482324]
9. US Department of Health and Human Services. Health Resources and Services Administration. 2014. Using Health Text Messages to Improve Consumer Health Knowledge, Behaviors, and Outcomes: An Environmental Scan URL: https://www.hrsa.gov/sites/default/files/archive/healthit/txt4tots/environmentalscan.pdf [accessed 2017-12-27] [WebCite Cache ID 6w13BtisT]
10. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
11. Kallander K, Tibenderana JK, Akpogheneta OJ, Strachan DL, Hill Z, ten Asbroek AHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low and middle income countries: a review. J Med Internet Res 2013;15(1):e17. [doi: 10.2196/jmir.2130]
12. Bull S, Ezeanochie N. From Foucault to Freire through Facebook: toward an integrated theory of mHealth. Health Educ Behav 2016 Aug;43(4):399-411. [doi: 10.1177/1090198115605310]
13. White A, Thomas DS, Ezeanochie N, Bull S. Health worker mHealth utilization: a systematic review. Comput Inform Nurs 2016 May;34(5):206-213. [doi: 10.1097/CIN.0000000000000231]
14. Hurt K, Walker RJ, Campbell JA, Egede LE. mHealth interventions in low and middle-income countries: a systematic review. Glob J Health Sci 2016 Sep 1;8(9):54429. [doi: 10.5539/gjhs.v8n9p183]
15. Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N. Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. PLoS One 2009;4(5):e5483 [FREE Full text] [doi: 10.1371/journal.pone.0005483] [Medline: 19434234]

Abbreviations

ANC: antenatal care
AR: autoregressive
ARV: antiretroviral
GEE: generalized estimating equation
GIS: geographic information system
LMIC: low and middle-income countries
MCH: maternal and child health
mHealth: mobile health
PMTCT: prevention of mother-to-child transmission
T-HIT: Tanzania Health Information Technology
WHO: World Health Organization

Edited by C Dias; submitted 19.07.17; peer-reviewed by M Peskin, M Fitz-Randolph, J Villegas; comments to author 24.08.17; revised version received 01.11.17; accepted 16.11.17; published 15.01.18.

Please cite as:
Bull S, Thomas DSK, Nyanza EC, Ngallaba SE
Tanzania Health Information Technology (T-HIT) System: Pilot Test of a Tablet-Based System to Improve Prevention of Mother-to-Child Transmission of HIV
JMIR Mhealth Uhealth 2018;6(1):e16
URL: https://mhealth.jmir.org/2018/1/e16/
doi:10.2196/mhealth.8513
PMID:29335236

Original Paper

Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry

Marcin Masalski[1,2], MD, PhD (Eng); Tomasz Grysiński[2], PhD (Eng); Tomasz Kręcicki[1], PhD, MD

[1]Department and Clinic of Otolaryngology, Head and Neck Surgery, Faculty of Postgraduate Medical Training, Wroclaw Medical University, Wrocław, Poland

[2]Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wroclaw University of Science and Technology, Wrocław, Poland

Corresponding Author:
Marcin Masalski, MD, PhD (Eng)
Department and Clinic of Otolaryngology, Head and Neck Surgery
Faculty of Postgraduate Medical Training
Wroclaw Medical University
Borowska 213
Wrocław, 50-556
Poland
Phone: 48 515086252
Email: marcin.masalski@pwr.edu.pl

Abstract

Background: Hearing screening tests based on pure-tone audiometry may be conducted on mobile devices, provided that the devices are specially calibrated for the purpose. Calibration consists of determining the reference sound level and can be performed in relation to the hearing threshold of normal-hearing persons. In the case of devices provided by the manufacturer, together with bundled headphones, the reference sound level can be calculated once for all devices of the same model.

Objective: This study aimed to compare the hearing threshold measured by a mobile device that was calibrated using a model-specific, biologically determined reference sound level with the hearing threshold obtained in pure-tone audiometry.

Methods: Trial participants were recruited offline using face-to-face prompting from among Otolaryngology Clinic patients, who own Android-based mobile devices with bundled headphones. The hearing threshold was obtained on a mobile device by means of an open access app, Hearing Test, with incorporated model-specific reference sound levels. These reference sound levels were previously determined in uncontrolled conditions in relation to the hearing threshold of normal-hearing persons. An audiologist-assisted self-measurement was conducted by the participants in a sound booth, and it involved determining the lowest audible sound generated by the device within the frequency range of 250 Hz to 8 kHz. The results were compared with pure-tone audiometry.

Results: A total of 70 subjects, 34 men and 36 women, aged 18-71 years (mean 36, standard deviation [SD] 11) participated in the trial. The hearing threshold obtained on mobile devices was significantly different from the one determined by pure-tone audiometry with a mean difference of 2.6 dB (95% CI 2.0-3.1) and SD of 8.3 dB (95% CI 7.9-8.7). The number of differences not greater than 10 dB reached 89% (95% CI 88-91), whereas the mean absolute difference was obtained at 6.5 dB (95% CI 6.2-6.9). Sensitivity and specificity for a mobile-based screening method were calculated at 98% (95% CI 93-100.0) and 79% (95% CI 71-87), respectively.

Conclusions: The method of hearing self-test carried out on mobile devices with bundled headphones demonstrates high compatibility with pure-tone audiometry, which confirms its potential application in hearing monitoring, screening tests, or epidemiological examinations on a large scale.

(JMIR Mhealth Uhealth 2018;6(1):e10) doi:10.2196/mhealth.7800

KEYWORDS

hearing test; mobile health; mobile apps; pure-tone audiometry

Introduction

This study investigated the accuracy of the hearing tests conducted on mobile devices calibrated by means of the biological method, that is, in relation to the hearing threshold of a normal-hearing person. This study constitutes the second part of the planned research. The calibration method was adopted from a previous study [1], and it involved semiautomated determination of predefined, model-specific reference sound level.

Hearing loss is a disorder widely encountered among the world population. It is estimated that it affects 5.3% of population, which totals 360 million patients suffering from hearing loss around the world [2]. Monitoring and diagnosis of hearing are especially important in preventing and treatment of hearing loss. One of the basic hearing tests is pure-tone audiometry, which determines the hearing threshold in relation to the sound frequency. However, pure-tone audiometry requires access to specialized medical equipment and staff.

Automated audiometry involves self-determining the hearing threshold and is a valuable diagnostic tool in the case of limited access to medical personnel [3-8]. The methods of automated audiometry have been developed over many years. At present, common methods involve the assessment of the reaction to an automatically generated test signal [4,8-21] or determining the lowest audible sound via self-adjustment of test signal intensity [6,7,9,22]. The hearing thresholds obtained by means of both methods on calibrated devices are comparable with conventional pure-tone audiometry [4-7,9,11-14,19,20,22].

Audiometric tests can be conducted on specialized equipment as well as on generally accessible electronic devices such as personal computers [10,22,23] or mobile devices [4-9,15-21,24,25]. However, reliable results may only be obtained if they are correctly calibrated and conducted in silence [1,15,21,26,27]. Assuming that mobile devices sold with bundled headphones have similar frequency characteristics of the sound system within the devices of the same model, it is possible to share calibration coefficients within these groups. Examinations on iOS-based devices confirmed the validity of automated audiometry based on common calibration coefficients determined in laboratory conditions for all devices of this type [4].

Besides applications typical for automated audiometry such as preliminary evaluation of audiological patients and screening tests [3-8], tests on common electronic devices may be useful for self-monitoring of hearing, especially during recovery from sudden sensorineural hearing loss; in Ménière's disease, tinnitus or other fluctuating hearing loss, or age-related hearing loss; or during ototoxic therapy [5,6,8,10,16,17,22]. The development of audiometric apps is stimulated by a wide range of potential applications. There is a large number of mobile apps on the market, but only few of them have been investigated in research, which is crucial before they can be clinically applied [28,29].

The objective of this trial was to determine the accuracy of automated audiometry conducted on Android-based mobile sets including the device and the bundled headphones that were calibrated semiautomatically via a biological method. Android-based devices constitute 80% of all mobile devices around the world, and their number is estimated to be 1.4 billion [30]. The prevalence of these devices may contribute to improving accessibility of audiological examinations, especially in the parts of the world with limited access to specialized equipment [5,7,24,31,32]. Android-based mobile devices, contrary to iOS-based ones, are produced by many manufacturers, which means that they are not unified in terms of hardware solutions. This is reflected by statistically significant differences that were found in frequency characteristic of the sound system [1]. Additionally, the number of Android-based models or their variations in 2015 exceeded 24,000, with an increase of 5000 compared with the previous year [33]. Due to the diversity of this group, its size, and dynamics of its substantial changes, laboratory calibration appears to be an inefficient solution.

Calibration of mobile devices conducted by means of a biological method involves determining reference sound level in relation to the hearing threshold of normal-hearing persons [1,34]. Calibration conducted several times on different mobile sets of the same model allows for determining a reliable, model-specific reference sound level [1]. The adopted solution provides support without additional workload not only for the current but also for the future models of mobile sets.

Methods

This study involved a comparison of hearing thresholds determined by pure-tone audiometry with the hearing thresholds determined on Android-based mobile devices calibrated biologically in uncontrolled conditions. It was a single-center, crossover trial carried out on patients of Otolaryngology Clinic. The consent to conduct the trial has been granted by the Bioethics Committee of Wroclaw Medical University.

Recruitment

Study participants were recruited offline, using face-to-face prompting. Eligibility assessment was conducted among subjects who owned an Android-based device with bundled headphones. Eligibility criterion, apart from the willingness to participate in the study, included owning the device, for which model-specific calibration coefficient has been previously determined by means of biological method [1]. All the study participants conducted examinations on their mobile devices.

Measurements

Measurements involved conducting pure-tone audiometry and the test on a mobile device. The order was arranged in a counterbalanced manner. The measurements for both methods were conducted in a sound booth for the frequencies 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, 6 kHz, and 8 kHz.

Pure-tone audiometry was conducted by means of conventional 10 dB down and 5 dB up bracketing method (modified Hughson-Westlake method) in accordance with the recommendations of the British Society of Audiology [35]. The measurements were conducted by an audiologist on a clinical audiometer Interacoustic AD229e using TDH-39 headphones

(Interacoustics, Denmark, Audiometer Allé, 5500 Middelfart) previously calibrated according to ISO 389-1:1998.

Hearing examinations on mobile devices were conducted using a free app Hearing Test [36,37] available on Google Play. The examinations involved self-determining of the lowest audible sound generated by the device and were audiologist-assisted. The participant changed intensity of the test sound using the buttons “I can hear” and “I cannot hear” and then confirmed the lowest audible sound using the button “Barely audible.” During the adjustment of the sound intensity, temporary test result was continuously presented on the device screen (Figure 1). The role of the supervising audiologist was to prevent obvious mistakes, such as switching the sides of headphones or omitting the frequency by accidental button pressing. The audiologist also provided assistance in cases of doubts concerning proper execution of the test. Test tone was amplitude modulated with the depth of 100% and the frequency of 2 Hz. Sound intensity was changed in 5 dB steps. When the intensity of the test sound exceeded 40 dB HL, a masking sound was generated contralaterally. The tests on mobile devices were conducted twice: test and retest. Before the retest, the headphones were removed and put on again.

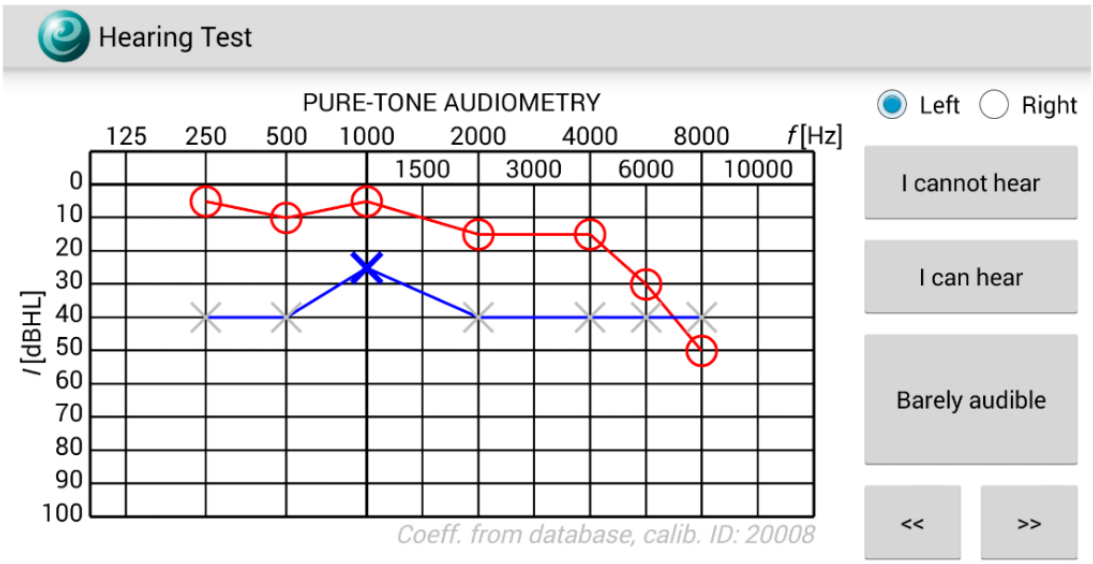
Calibration of mobile devices was described in detail in the previous paper [1]. The objective of calibration was to determine the intensity of the signal generated by the device, which would generate a sound in bundled headphones at the reference level of 0 dB HL. Calibration was performed by means of a biological method, which led to elaboration of a new mobile hearing level (mHL) scale. The level of 0 dB mHL was defined as the sound level generated by the mobile set equivalent to the hearing threshold of normal-hearing persons, reduced by their estimated hearing threshold. The advantage of this approach is the ability to calibrate new device models based on measurements conducted by users themselves, which is essential due to high rotation of mobile models on the market. The calibration is required only for users who are the first to become the owners of new devices, whereas the others can reuse their results.

Calibration measurements involved self-determining the hearing thresholds by means of Bekesy’s method [1,10,11,22,34] by users aged 18-35 years who considered themselves normal-hearing persons. There were no other requirements for calibration. However, if these requirements were not met, a user could ask another person for help. These calibration thresholds were then decreased by the population median to obtain the final reference level. More precisely, the estimator of central tendency of the determined thresholds were approximated by 37th percentile and then decreased by the median of the hearing threshold estimated based on literature data [38] for the corresponding population, that is, aged 18-35 years [1]. Calibration measurements were conducted in uncontrolled conditions, irrespective of this research, before its commencement and when it was in progress, by application users who were not trial participants. To calibrate the model, at least 15 measurements were required to be conducted on various devices belonging to the same model [1]. After each subsequent measurement, the calibration results were automatically updated.

Statistical Analysis

Sample size was determined on the basis of the standard deviation (SD) between the hearing thresholds in pure-tone audiometry and thresholds determined on mobile devices estimated at 8.42 dB [1]. With the statistical significance of 0.05, statistical power of 0.8, and the effect size of 2.0 dB, the sample volume was estimated on 140 ears (70 subjects). Hearing thresholds obtained on mobile devices have been compared with pure-tone audiometry and analyzed in test-retest examination. Differences in hearing thresholds, intraclass correlation coefficients, and Cronbach alpha with respective confidence intervals have been determined. The results were presented in dependency plots as well as in Bland-Altman plots.

Figure 1. Screenshot of the Hearing Test app during examination.



Results

In the period between November 11, 2015, and November 2, 2016, 89 subjects were assessed for eligibility from among patients of Otolaryngology Department who owned a mobile device with bundled headphones. Out of 89 subjects, 7 (8%) were not interested in the examination and did not give their consent; 12 out of the remaining 82 subjects (15%) owned devices that were not yet calibrated (Figure 2).

A total of 70 subjects, including 34 men and 36 women, aged 18-71 years (mean 36, SD 11) participated in the trial. Among the participants were both normal-hearing subjects and hearing-impaired patients. Table 1 presents the pure-tone audiometry thresholds for tested frequencies, Table 2 shows distribution of the hearing loss by type, and Table 3 summarizes the models of mobile devices applied.

The research was arranged in a counterbalanced manner. Out of 70 subjects, 35 (50%) were first tested by pure-tone audiometry, whereas the other half by the mobile device. Out of the range measurements were discarded from further analysis: 13 of 980 (1.3%), 14 of 966 (1.4%), and 10 of 980 (1.0%) in the case of test, retest, and pure-tone audiometry, respectively. Hearing thresholds determined through pure-tone audiometry were compared with thresholds obtained on mobile devices separately for both groups (Table 4). At the level of statistical significance of P=.05, no differences were found between groups, and thus, further analyses were conducted jointly (Tables 5 and 6, and Figures 3 and 4). The mean difference between the hearing threshold determined by pure-tone audiometry and on mobile devices was significantly different from 0 and reached 2.6 dB (95% CI 2.0-3.1), with SD of 8.3 dB (95% CI 7.9-8.7). Intraclass correlation for consistency of single measurement was obtained at the level of 0.85 (95% CI 0.83-0.87), whereas the Cronbach alpha for reliability was calculated at 0.92 (95% CI 0.91-0.93). The number of within-subject differences not greater than 10 dB reached 89% (95% CI 88-91), whereas the mean absolute difference was 6.5 dB (95% CI 6.2-6.9). The largest differences were noted for the frequencies of 6 and 8 kHz at the level of 4.0 dB (95% CI 2.5-5.4) and 7.0 dB (95% CI 5.4-8.6), respectively, whereas the lowest, statistically insignificant at P=.05, for the frequency of 500 Hz and 1 kHz.

Figure 2. Flow diagram.

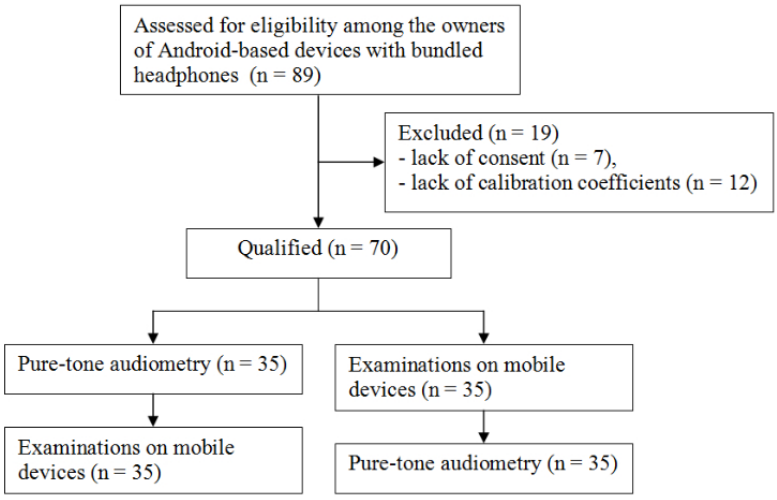


Table 1. Pure-tone audiometry thresholds among study participants.

Hearing loss (dB HL)	Number of ears						
	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	6 kHz	8 kHz
15 and less	108	108	118	118	110	87	93
16-25	14	16	11	12	12	22	15
26-40	11	11	8	8	7	15	13
41-60	6	4	2	0	4	7	8
60-80	0	0	0	1	5	5	5
Over 80	1	1	1	1	2	4	6

Table 2. Type of hearing loss among study participants.

Type of hearing loss	Number of participants
Normal hearing[a]	36
Sensorineural hearing loss	
Sudden deafness	6
Presbyaccusis	3
Neuronitis vestibularis	3
Postviral complication	1
Other and indefinite sensorineural hearing loss	8
Conductive and mixed hearing loss	
Chronic otitis	9
Cholesteatoma	2
Exudative otitis	1
Other and indefinite conductive or mixed hearing loss	1

[a]Hearing threshold in pure-tone audiometry no higher than 25 dB HL for all tested frequencies.

Table 3. Mobile devices by models.

Manufacturer and model	Number of devices
Samsung SM-G900F	7
Huawei ALE-L21	4
Samsung SM-G350	4
Samsung GT-I9300	3
Samsung GT-I9505	3
Samsung SM-A500FU	3
Samsung SM-G386F	3
Samsung SM-G530FZ	3
Samsung GT-I8190N	2
Samsung GT-I9060I	2
Samsung GT-I9515	2
Samsung SM-G800F	2
Samsung SM-A300FU	2
Sony D6603	2
Sony D2303	2
Sony E2303	2
Other	24

Table 4. Differences in hearing threshold determined by pure-tone audiometry and on mobile devices depending on the order of examinations.

Frequency in Hz	First pure-tone audiometry		First examination on a mobile device	
	n	Mean difference in dB (95% CI)	n	Mean difference in dB (95% CI)
250	70	3.6 (1.3-5.8)	69	1.8 (−0.2 to 3.8)
500	70	−0.8 (−2.6 to 1.0)	69	−0.4 (−2.5 to 1.6)
1000	70	−0.4 (−2.1 to 1.2)	69	−1.7 (−3.1 to −0.2)
2000	70	2.6 (1.2-3.9)	69	3.3 (1.9-4.7)
4000	70	2.4 (0.4-4.3)	69	4.1 (2.4-5.9)
6000	70	4.0 (2.2-5.8)	69	3.9 (1.6-6.2)
8000	69	6.5 (4.2-8.8)	64	7.5 (5.3-9.7)
Total	489	2.5 (1.8-3.3)	478	2.6 (1.9-3.4)

Table 5. Comparison of the hearing threshold determined by pure-tone audiometry and on mobile devices.

Frequency in Hz	n	Mean difference in dB (95% CI)	Standard deviation in dB (95% CI)	Mean absolute difference in dB (95% CI)	Intraclass correlation[a] (95% CI)
250	139	2.7 (1.2-4.2)	9.0 (8.1-10.2)	6.9 (6.0-7.9)	0.75 (0.67-0.82)
500	139	−0.6 (−2.0 to 0.7)	8.0 (7.2-9.1)	5.9 (5.2-6.8)	0.79 (0.72-0.85)
1000	139	−1.0 (−2.1 to 0.1)	6.6 (5.9-7.4)	4.8 (4.1-5.5)	0.80 (0.73-0.85)
2000	139	2.9 (2.0-3.9)	5.7 (5.1-6.4)	4.9 (4.3-5.5)	0.86 (0.82-0.90)
4000	139	3.2 (1.9-4.5)	7.8 (7.0-8.8)	6.7 (5.9-7.6)	0.90 (0.86-0.92)
6000	139	4.0 (2.5-5.4)	8.4 (7.5-9.6)	7.3 (6.4-8.2)	0.90 (0.87-0.93)
8000	133	7.0 (5.4-8.6)	9.3 (8.3-10.6)	9.3 (8.2-10.5)	0.85 (0.80-0.89)
Total	967	2.6 (2.0-3.1)	8.3 (7.9-8.7)	6.5 (6.2-6.9)	0.85 (0.83-0.87)

[a]Two-way random for consistency of single measurement.

Table 6. Percentage distribution of within-subject differences in the hearing threshold determined by pure-tone audiometry and on mobile devices.

Frequency (Hz)	n	0-5 dB (%)	0-10 dB (%)	0-15 dB (%)	0-20 dB (%)	0-25 dB (%)	0-30 dB (%)
250	139	74	87	92	96	99	100
500	139	87	94	98	100	100	100
1000	139	93	97	98	99	100	100
2000	139	81	96	100	100	100	100
4000	139	69	91	97	99	100	100
6000	139	66	86	94	99	100	100
8000	133	50	74	89	97	98	100
Total	967	74	89	95	99	100	100

Figure 3. Hearing threshold determined on mobile devices (mHT) in relation to the hearing threshold in pure-tone audiometry (HT).

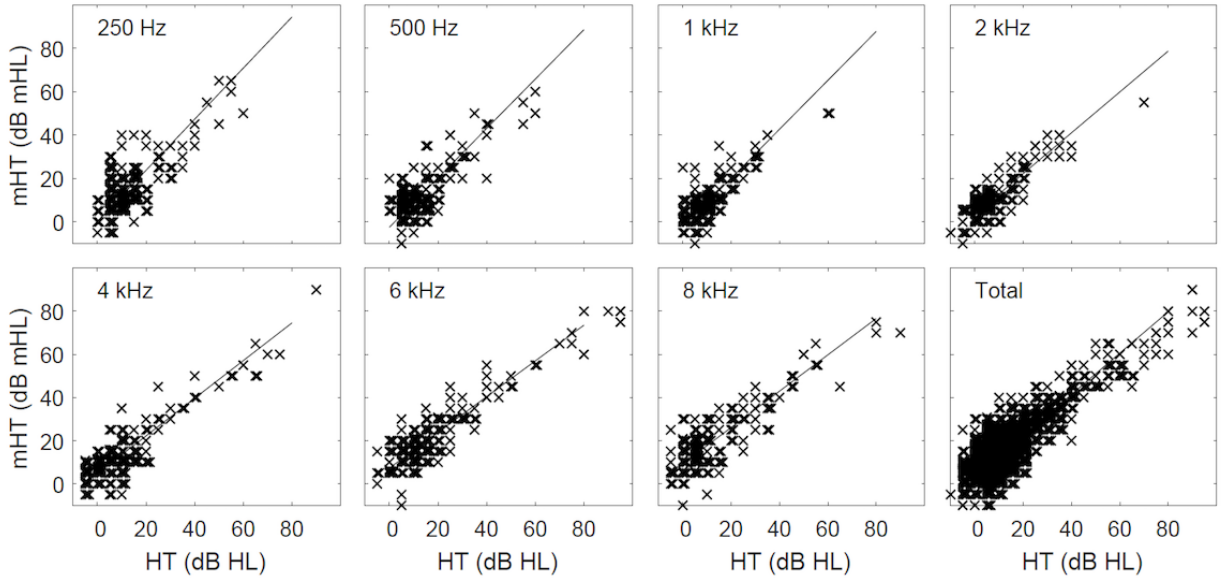
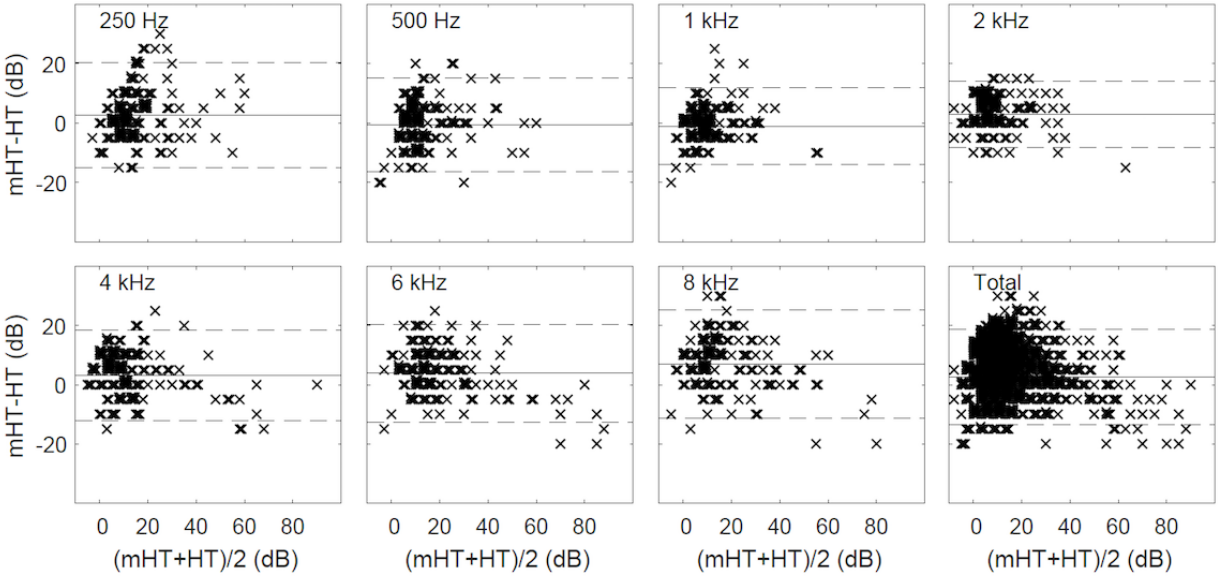


Figure 4. Difference plots (Bland-Altman) between the hearing threshold determined by pure-tone audiometry (HT) and on mobile devices (mHT) (continuous line indicates mean difference; dashed line indicates ±1.96 standard deviation range).



Retest was conducted on a mobile device in 69 of 70 subjects (99%). At the level of statistical significance of P=.05, no differences in hearing threshold were found at any frequency (Table 7). The mean difference was −0.1 dB (95% CI −0.4 to 0.2) with SD of 4.4 dB (95% CI 4.2-4.6); 99% (95% CI 99-100) of within-subject differences were within the range of 10 dB, and the mean absolute difference reached 2.8 dB (95% CI 2.6-3.0) (Tables 7 and 8).

The assessment of specificity and sensitivity of hearing loss detection was conducted based on the criterion [39] adopted for the purposes of this study. Hearing loss was diagnosed when the threshold exceeded 30 dB at one of the following frequencies: 500 Hz, 1 kHz, 2 kHz, or 25 dB at more than one, or when the hearing threshold exceeded 50 dB at 4 kHz. The obtained sensitivity was at 98% (95% CI 93-100) with specificity of 79% (95% CI 71-87).

Table 7. Comparison of hearing thresholds determined on mobile devices in test-retest examination.

Frequency in Hz	n	Mean difference in dB (95% CI)	Standard deviation in dB (95% CI)	Mean absolute difference in dB (95% CI)	Intraclass correlation[a] (95% CI)
250	137	0.0 (−0.7 to 0.7)	4.4 (3.9-5.0)	2.9 (2.5-3.4)	0.95 (0.93-0.96)
500	137	−0.3 (−1.0 to 0.5)	4.3 (3.8-4.9)	2.7 (2.3-3.2)	0.94 (0.92-0.96)
1000	137	0.1 (−0.6 to 0.8)	4.3 (3.9-4.9)	2.6 (2.2-3.0)	0.92 (0.89-0.94)
2000	137	−0.4 (−1.0 to 0.2)	3.6 (3.3-4.1)	2.4 (2.0-2.8)	0.94 (0.92-0.96)
4000	137	0.0 (−0.7 to 0.7)	4.2 (3.8-4.8)	2.7 (2.3-3.1)	0.97 (0.95-0.98)
6000	137	−0.1 (−0.9 to 0.7)	4.5 (4.1-5.2)	3.2 (2.8-3.7)	0.97 (0.95-0.98)
8000	130	−0.2 (−1.1 to 0.7)	5.1 (4.6-5.9)	3.2 (2.6-3.7)	0.95 (0.93-0.96)
Total	952	−0.1 (−0.4 to 0.2)	4.4 (4.2-4.6)	2.8 (2.6-3.0)	0.96 (0.95-0.96)

[a]Two-way random for consistency of single measurement.

Table 8. Percentage distribution of within-subject differences in the hearing threshold determined on mobile devices in test-retest examination.

Frequency (Hz)	n	0-5 dB (%)	0-10 dB (%)	0-15 dB (%)
250	137	97	99	100
500	137	98	99	100
1000	137	96	99	100
2000	137	99	100	100
4000	137	97	100	100
6000	137	97	100	100
8000	130	97	98	100
Total	952	97	99	100

Discussion

Principal Findings

In this paper, hearing thresholds determined by pure-tone audiometry were compared with hearing thresholds determined on Android-based mobile devices previously calibrated biologically in uncontrolled conditions. At the level of statistical significance P=.05, no differences in relation to the test order were observed. The mean hearing threshold difference of 2.6 dB (SD 8.3 dB) confirms the reliability of the method. Additionally, the mean test-retest difference of −0.1 dB (SD 4.4) indicates its high repeatability.

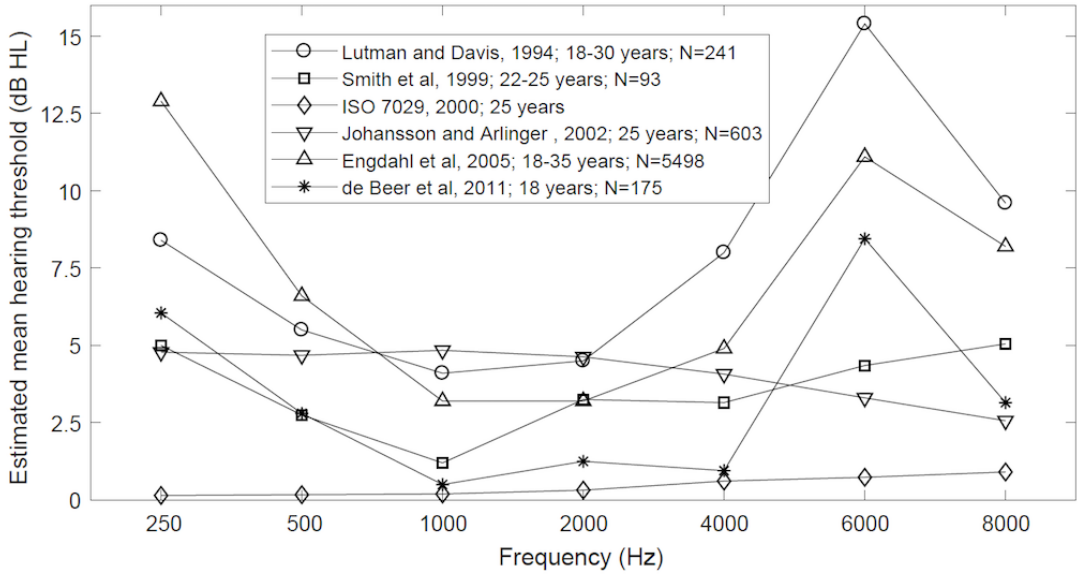
Comparison With Prior Study

The results are consistent with the research presented in previous works. The SD of the difference at 8.3 dB (95% CI 7.9-8.7) corresponds to the value estimated for this calibration method of 8.42 dB [1] and is close to the value of 7.8 dB obtained in another study conducted on the same app Hearing Test [7]. The mean difference of 2.6 dB (95% CI 2.0-3.1) and the mean absolute difference of 6.5 dB (95% CI 6.2-6.9) are also comparable with those presented in [7], that is, 0.7 dB and 7.8 dB, respectively. Contrary to study [7], in which all the tests were carried out on one device, each test in this trial was performed on a different device, thus validating this calibration method.

SD of the difference for the proposed method was 8.3 dB (95% CI 7.9-8.7) with the number of differences no higher than 10 dB at 89% (95% CI 88-91). These values are comparable with the results determined on a set calibrated by a normal-hearing person with verified hearing threshold (SD 6.9 and 8.29 dB [10] and SD 7.88 dB [22]) and are significantly smaller when based on a single biological calibration conducted in uncontrolled conditions (SD 10.66 dB [22]). The highest calibration accuracy is obtained in laboratory conditions, which, however, is not scalable for the Android-based devices. In laboratory conditions, SD varies from 6.4 dB to 9.9 dB [5,6,25], whereas the number of differences not higher than 10 dB changes from 81% up to 96% [4,5,6,19], depending on other settings.

SD of test-retest difference was obtained at 4.4 dB (95% CI 4.2-4.6) with the number of differences no greater than 10 dB at 99% (95% CI 99-100). The results are consistent with other works (SD 4.97 [34], number of differences 97% [6]) using the method of determining the lowest audible sound through self-adjustment of the intensity of the test signal. Moreover, these values do not differ much from the test-retest examinations obtained for automated 10 dB down and 5 dB up bracketing method (SD 6.05 dB and 5.00 dB [34], the number of differences 98% [13]). This confirms the comparable accuracy of both methods [9], at least in the cases of mobile device owners who are familiar with the use of touch screens. Additionally, the results of test-retest examinations confirm the efficiency of the method, especially in self-monitoring of hearing.

Figure 5. Mean hearing thresholds in a population of young normal-hearing adults estimated on the basis of literature data.



Limitations

Tests were audiologist-assisted and conducted in a sound booth; therefore, the results are devoid of user mistakes such as omitting the frequency by accidental button pressing or switching the sides of headphones as well as performing the examination in noise. When tested in uncontrolled conditions, care should be taken to minimize the risk of such mistakes, for example, by monitoring the test duration at single frequency and tracking the background noise using a built-in microphone.

The factors limiting the accessibility of the test in relation to the number of Android-based mobile devices are bundled headphones and calibration coefficients. Not all the devices are offered with bundled headphones. Moreover, calibration coefficients for less popular devices may not be determined if the required number of calibrations is not achieved.

During tests on mobile devices, temporary result was continuously displayed on the screen. Therefore, during the retest, subjects could be biased by previous results, consequently reducing the test-retest difference.

Mean difference in the hearing threshold determined through pure-tone audiometry and on mobile devices differed significantly from 0 dB and reached 7.0 dB (95% CI 5.5-8.5) at 8 kHz. The reason for such discrepancies may be related to the literature-based median values of the hearing threshold of normal-hearing subjects, which were adopted to determine 0 dB HL. The mean hearing threshold estimated on the basis of literature data for people aged between 18 and 35 years displays considerable differences (Figure 5) [38,40-44]. It may be related to differing definitions of the normal-hearing person, standards of the measurement method used, or other factors such as genetic composition or occupational distribution of the population. Therefore, the determination of the values for adoption is complex. This research applies values presented in study [38]. In the future, however, it seems rational to correct them by taking into consideration the identified discrepancies (Table 5).

Conclusions

The method of hearing self-test carried out on mobile devices with bundled headphones calibrated by model-specific coefficients determined in relation to the hearing thresholds of normal-hearing persons demonstrates high compatibility with pure-tone audiometry, which confirms its potential application in hearing monitoring, screening tests, or epidemiological examinations on a large scale. However, one must acknowledge its limitations resulting mainly from the calibration method. Further evaluation of the efficiency of the method, especially in particular applications, is justified and required.

Acknowledgments

The trial was cofinanced by Wroclaw Medical University (Research Grant for Young Scientists, Grant No 170/Pbmn and the statutory funds) as well as Wrocław University of Science and Technology (the statutory funds).

Conflicts of Interest

MM is the author and owner of the Hearing Test app, by means of which the measurements described in this paper were conducted.

References

1. Masalski M, Kipiński L, Grysiński T, Kręcicki T. Hearing tests on mobile devices: evaluation of the reference sound level by means of biological calibration. J Med Internet Res 2016 May 30;18(5):e130 [FREE Full text] [doi: 10.2196/jmir.4987] [Medline: 27241793]
2. World Health Organization. Geneva, Switzerland; 2017 Feb. Deafness and hearing loss URL: http://www.who.int/mediacentre/factsheets/fs300/en/ [accessed 2017-12-14] [WebCite Cache ID 6oIIlrNL0]
3. Margolis RH, Morgan DE. Automated pure-tone audiometry: an analysis of capacity, need, and benefit. Am J Audiol 2008 Dec;17(2):109-113. [doi: 10.1044/1059-0889(2008/07-0047)] [Medline: 18840703]
4. Foulad A, Bui P, Djalilian H. Automated audiometry using apple iOS-based application technology. Otolaryngol Head Neck Surg 2013 Nov;149(5):700-706. [doi: 10.1177/0194599813501461] [Medline: 23963611]
5. Sandström J, Swanepoel DW, Carel Myburgh H, Laurent C. Smartphone threshold audiometry in underserved primary health-care contexts. Int J Audiol 2016;55(4):232-238. [doi: 10.3109/14992027.2015.1124294] [Medline: 26795898]
6. Kam AC, Sung JK, Lee T, Wong TK, van Hasselt A. Clinical evaluation of a computerized self-administered hearing test. Int J Audiol 2012 Aug;51(8):606-610. [doi: 10.3109/14992027.2012.688144] [Medline: 22676699]
7. Renda L, Selçuk Ö, Eyigör H, Osma Ü, Yılmaz MD. Smartphone based audiometric test for confirming the level of hearing; is it useable in underserved areas? J Int Adv Otol 2016 Apr;12(1):61-66 [FREE Full text] [doi: 10.5152/iao.2016.1421] [Medline: 27340985]
8. Peer S, Fagan JJ. Hearing loss in the developing world: evaluating the iPhone mobile device as a screening tool. S Afr Med J 2015 Jan;105(1):35-39. [Medline: 26046161]
9. Van Tasell DJ, Folkeard P. Reliability and accuracy of a method of adjustment for self-measurement of auditory thresholds. Otol Neurotol 2013 Jan;34(1):9-15. [doi: 10.1097/MAO.0b013e318278c05d] [Medline: 23202154]
10. Honeth L, Bexelius C, Eriksson M, Sandin S, Litton J, Rosenhall U, et al. An internet-based hearing test for simple audiometry in nonclinical settings: preliminary validation and proof of principle. Otol Neurotol 2010 Jul;31(5):708-714. [doi: 10.1097/MAO.0b013e3181de467a] [Medline: 20458255]
11. McPherson B, Law MM, Wong MS. Hearing screening for school children: comparison of low-cost, computer-based and conventional audiometry. Child Care Health Dev 2010 May;36(3):323-331. [doi: 10.1111/j.1365-2214.2010.01079.x] [Medline: 20507326]
12. Margolis RH, Glasberg BR, Creeke S, Moore BCJ. AMTAS: automated method for testing auditory sensitivity: validation studies. Int J Audiol 2010 Mar;49(3):185-194. [doi: 10.3109/14992020903092608] [Medline: 20109081]
13. Swanepoel DW, Mngemane S, Molemong S, Mkwanazi H, Tutshini S. Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. Telemed J E Health 2010 Jun;16(5):557-563. [doi: 10.1089/tmj.2009.0143] [Medline: 20575723]
14. Ho ATP, Hildreth AJ, Lindsey L. Computer-assisted audiometry versus manual audiometry. Otol Neurotol 2009 Oct;30(7):876-883. [doi: 10.1097/MAO.0b013e3181b120d0] [Medline: 20179426]
15. Khoza-Shangase K, Kassner L. Automated screening audiometry in the digital age: exploring uhear™ and its use in a resource-stricken developing country. Int J Technol Assess Health Care 2013 Jan;29(1):42-47. [doi: 10.1017/S0266462312000761] [Medline: 23298579]
16. Handzel O, Ben-Ari O, Damian D, Priel MM, Cohen J, Himmelfarb M. Smartphone-based hearing test as an aid in the initial evaluation of unilateral sudden sensorineural hearing loss. Audiol Neurootol 2013;18(4):201-207. [doi: 10.1159/000349913] [Medline: 23689282]
17. Abu-Ghanem S, Handzel O, Ness L, Ben-Artzi-Blima M, Fait-Ghelbendorf K, Himmelfarb M. Smartphone-based audiometric test for screening hearing loss in the elderly. Eur Arch Otorhinolaryngol 2015 Feb 6;273(2):333-339. [doi: 10.1007/s00405-015-3533-9] [Medline: 25655259]
18. Szudek J, Ostevik A, Dziegielewski P, Robinson-Anagor J, Gomaa N, Hodgetts B, et al. Can Uhear me now? validation of an iPod-based hearing loss screening test. J Otolaryngol Head Neck Surg 2012 Apr;41 Suppl 1:S78-S84. [Medline: 22569055]
19. Saliba J, Al-Reefi M, Carriere JS, Verma N, Provencal C, Rappaport JM. Accuracy of mobile-based audiometry in the evaluation of hearing loss in quiet and noisy environments. Otolaryngol Head Neck Surg 2017 Apr;156(4):706-711. [doi: 10.1177/0194599816683663] [Medline: 28025906]
20. Yeung J, Javidnia H, Heley S, Beauregard Y, Champagne S, Bromwich M. The new age of play audiometry: prospective validation testing of an iPad-based play audiometer. J Otolaryngol Head Neck Surg 2013;42:21 [FREE Full text] [doi: 10.1186/1916-0216-42-21] [Medline: 23663317]
21. Yeung JC, Heley S, Beauregard Y, Champagne S, Bromwich MA. Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. Int J Pediatr Otorhinolaryngol 2015 Aug;79(8):1248-1252. [doi: 10.1016/j.ijporl.2015.05.021] [Medline: 26055197]
22. Masalski M, Kręcicki T. Self-test web-based pure-tone audiometry: validity evaluation and measurement error analysis. J Med Internet Res 2013;15(4):e71 [FREE Full text] [doi: 10.2196/jmir.2222] [Medline: 23583917]
23. Choi JM, Lee HB, Park CS, Oh SH, Park KS. PC-based tele-audiometry. Telemed J E Health 2007 Oct;13(5):501-508. [doi: 10.1089/tmj.2007.0085] [Medline: 17999612]

24. Swanepoel DW, Myburgh HC, Howe DM, Mahomed F, Eikelboom RH. Smartphone hearing screening with integrated quality control and data management. Int J Audiol 2014 Dec;53(12):841-849. [doi: 10.3109/14992027.2014.920965] [Medline: 24998412]
25. Larrosa F, Rama-Lopez J, Benitez J, Morales JM, Martinez A, Alañon MA, et al. Development and evaluation of an audiology app for iPhone/iPad mobile devices. Acta Otolaryngol 2015;135(11):1119-1127. [doi: 10.3109/00016489.2015.1063786] [Medline: 26144548]
26. Al-Abri R, Al-Balushi M, Kolethekkat A, Bhargava D, Al-Alwi A, Al-Bahlani H, et al. The accuracy of IOS device-based uHear as a screening tool for hearing loss: a preliminary study from the Middle East. Oman Med J 2016 Mar;31(2):142-145 [FREE Full text] [doi: 10.5001/omj.2016.27] [Medline: 27168926]
27. Livshitz L, Ghanayim R, Kraus C, Farah R, Even-Tov E, Avraham Y, et al. Application-based hearing screening in the elderly population. Ann Otol Rhinol Laryngol 2017 Jan;126(1):36-41. [doi: 10.1177/0003489416672873] [Medline: 27913721]
28. Bright T, Pallawela D. Validated smartphone-based apps for ear and hearing assessments: a review. JMIR Rehabil Assist Technol 2016 Dec 23;3(2):e13. [doi: 10.2196/rehab.6074]
29. Paglialonga A, Tognola G, Pinciroli F. Apps for hearing science and care. Am J Audiol 2015 Sep;24(3):293-298. [doi: 10.1044/2015_AJA-14-0093] [Medline: 26649533]
30. Statista. Number of smartphone users worldwide from 2014 to 2020 URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2017-02-15] [WebCite Cache ID 6oIJ3U8fZ]
31. Swanepoel DW, Clark JL, Koekemoer D, Hall JW, Krumm M, Ferrari DV, et al. Telehealth in audiology: the need and potential to reach underserved communities. Int J Audiol 2010 Mar;49(3):195-202. [doi: 10.3109/14992020903470783] [Medline: 20151929]
32. Fagan JJ, Jacobs M. Survey of ENT services in Africa: need for a comprehensive intervention. Glob Health Action 2009;2 [FREE Full text] [doi: 10.3402/gha.v2i0.1932] [Medline: 20027268]
33. Open Signal. 2015 Aug. Android fragmentation visualized URL: https://opensignal.com/reports/2015/08/android-fragmentation/ [accessed 2017-12-14] [WebCite Cache ID 6oIJCbx8R]
34. Masalski M, Grysiński T, Kręcicki T. Biological calibration for web-based hearing tests: evaluation of the methods. J Med Internet Res 2014;16(1):e11 [FREE Full text] [doi: 10.2196/jmir.2798] [Medline: 24429353]
35. British Society of Audiology. Recommended procedure: pure-tone air-conduction and bone-conduction threshold audiometry with and without masking URL: http://www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf [accessed 2017-12-14] [WebCite Cache ID 6vi7XSH9S]
36. Masalski M. Google Play. Hearing Test URL: https://play.google.com/store/apps/details?id=mobile.eaudiologia&hl=en [accessed 2017-12-14] [WebCite Cache ID 6TZUfFaE9]
37. Amlani AM. Apps for the ears: tablet and smartphone users can choose from a growing number of apps that test hearing sensitivity and amplify sound. The ASHA Leader 2014 Jul;19:34-35 [FREE Full text] [doi: 10.1044/leader.APP.19072014.34]
38. Engdahl B, Tambs K, Borchgrevink HM, Hoffman HJ. Screened and unscreened hearing threshold levels for the adult population: results from the nord-trøndelag hearing loss study. Int J Audiol 2005 Apr;44(4):213-230. [Medline: 16011050]
39. Klockhoff I, Drettner B, Hagelin KW, Lindholm L. A method for computerized classification of pure tone screening audiometry results in noise-exposed groups. Acta Otolaryngol 1973 Apr;75(4):339-340. [Medline: 4702637]
40. Lutman ME, Davis AC. The distribution of hearing threshold levels in the general population aged 18-30 years. Audiology 1994;33(6):327-350. [Medline: 7741666]
41. Smith P, Davis A, Ferguson M, Lutman M. Hearing in young adults: report to ISO/TC43/WG1. Noise Health 1999;1(4):1-10 [FREE Full text] [Medline: 12689486]
42. ISO. 2000. Acoustics: statistical distribution of hearing thresholds as a function of age URL: https://www.iso.org/standard/26314.html [accessed 2017-12-22] [WebCite Cache ID 6vtT5E02l]
43. Johansson MSK, Arlinger SD. Hearing threshold levels for an otologically unscreened, non-occupationally noise-exposed population in Sweden. Int J Audiol 2002 Apr;41(3):180-194. [Medline: 12033636]
44. de Beer B, Graamans K, Zielhuis G, Ad S. Hearing thresholds of otologically healthy 18-year-olds. J Hear Sci 2011;1(2):27-34. [doi: 10.17430/882058]

Abbreviations

HT: hearing threshold determined by pure-tone audiometry
mHL: mobile hearing level
mHT: hearing threshold determined on mobile devices
SD: standard deviation

Edited by G Eysenbach; submitted 05.04.17; peer-reviewed by A Paglialonga, J Rey-Martinez; comments to author 12.07.17; revised version received 07.09.17; accepted 30.10.17; published 10.01.18.

Please cite as:
Masalski M, Grysiński T, Kręcicki T
Hearing Tests Based on Biologically Calibrated Mobile Devices: Comparison With Pure-Tone Audiometry
JMIR Mhealth Uhealth 2018;6(1):e10
URL: https://mhealth.jmir.org/2018/1/e10/
doi:10.2196/mhealth.7800
PMID:29321124

Original Paper

Translation and Validation of the Nomophobia Questionnaire in the Italian Language: Exploratory Factor Analysis

Mohammad Adawi[1*], MD, MHA; Nicola Luigi Bragazzi[2*], MD, PhD, MPH; Lidia Argumosa-Villar[3], BSc; Joan Boada-Grau[3], PhD; Andreu Vigil-Colet[3], PhD; Caglar Yildirim[4], PhD; Giovanni Del Puente[5], MD; Abdulla Watad[5,6], MD

[1]Bar-Ilan Faculty of Medicine, Padeh and Ziv Hospitals, Zafat, Israel

[2]Department of Health Sciences, Postgraduate School of Public Health, University of Genoa, Genoa, Italy

[3]Rovira i Virgili University, Tarragona, Spain

[4]Department of Computer Science, State University of New York at Oswego, Oswego, NY, United States

[5]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

[6]Zabludowicz Center for Autoimmune Diseases, Department of Medicine B, Sheba Medical Center, Tel Hashomer, Israel

[*]these authors contributed equally

Corresponding Author:
Nicola Luigi Bragazzi, MD, PhD, MPH
Department of Health Sciences
Postgraduate School of Public Health
University of Genoa
Via Antonio Pastore 1
Genoa, 16132
Italy
Phone: 39 0103538508
Fax: 39 0103538541
Email: robertobragazzi@gmail.com

Abstract

Background: Nomophobia, which is a neologism derived from the combination of “no mobile,” “phone,” and “phobia” is considered to be a modern situational phobia and indicates a fear of feeling disconnected.

Objective: No psychometric scales are available in Italian for investigating such a construct. We therefore planned a translation and validation study of the Nomophobia Questionnaire (NMP-Q), which is an instrument developed by Yildirim and Correia. Subjects were recruited via an online survey using a snowball approach.

Methods: The NMP-Q was translated from English into Italian using a classical “backwards and forwards” procedure. In order to explore the underlying factor structure of the translated questionnaire, an exploratory factor analysis was carried out. A principal component analysis approach with varimax rotation was performed. Multivariate regression analyses were computed to shed light on the psychological predictors of nomophobia.

Results: A sample of 403 subjects volunteered to take part in the study. The average age of participants was 27.91 years (standard deviation 8.63) and the sample was comprised of 160 males (160/403, 39.7%) and 243 females (243/403, 60.3%). Forty-five subjects spent less than 1 hour on their mobile phone per day (45/403, 11.2%), 94 spent between 1 and 2 hours (94/403, 23.3%), 69 spent between 2 and 3 hours (69/403, 17.1%), 58 spent between 3 and 4 hours (58/403, 14.4%), 48 spent between 4 and 5 hours (48/403, 11.9%), 29 spent between 5 and 7 hours (29/403, 7.2%), 36 spent between 7 and 9 hours (36/403, 8.9%), and 24 spent more than 10 hours (24/403, 6.0%). The eigenvalues and scree plot supported a 3-factorial nature of the translated questionnaire. The NMP-Q showed an overall Cronbach alpha coefficient of 0.95 (0.94, 0.89, and 0.88 for the three factors). The first factor explained up to 23.32% of the total variance, while the second and third factors explained up to 23.91% and 18.67% of the variance, respectively. The total NMP-Q score correlated with the number of hours spent on a mobile phone.

Conclusions: The Italian version of the NMP-Q proved to be reliable.

(JMIR Mhealth Uhealth 2018;6(1):e24) doi:10.2196/mhealth.9186

KEYWORDS

nomophobia; questionnaire validation; exploratory factor analysis; psychometric properties

Introduction

New information and communication technologies (ICTs) have become particularly widespread and are increasingly utilized in modernized cultures. Due to their frequent use and their ubiquitous and ever-present nature, ICTs are often perceived as an irreplaceable part of a highly dynamic and interconnected society [1]. In particular, smartphones, as the latest evolution of mobile ICTs [2], have signaled the start of the *mobile age*. According to the 2015 Pew Research Center survey, in 2015 approximately two-thirds of Americans had a smartphone, and this figure is expected to grow in the coming years [3]. In Italy, according to the 2015 survey carried out by the National Institute of Statistics, 54% of families had a smartphone in 2014, which represents an upward trend with respect to 2013 [4].

However, although mobile devices enable users to perform a variety of tasks in an unprecedentedly rapid, easy, and effective way, they can also lead to serious medical problems. These problems include exposure to radiation, "screen dermatitis," tumors, and infertility [5,6]; mobile devices can also interfere with driving safety and cause accidents [7,8]. Furthermore, the pervasiveness and intrusiveness of such devices could lead to mental health issues such as problematic use, distress, and compulsive usage, culminating in what is termed as "smartphone dependence" or "smartphone addiction" [9]. Some scholars consider this behavioral addiction to be a variant of technology addiction (or technopathy), while others consider it to be a specific addiction. Experts describe smartphone-related psychological disorders and syndromes, such as "short message service texting addiction," compulsive selfie-taking behavior, *sexting*, or *phubbing*, among others [10]. *Sexting* can be defined as a behavior consisting of sending and receiving sexually explicit messages [11], whereas *phubbing* (snubbing someone in a social setting in favor of one's own mobile phone) may lead to the impairment of social life and couple relationships [12].

Nomophobia, a neologism that is derived from the combination of "no mobile," "phone," and "phobia" has recently emerged as a modern problem, denoting the fear of feeling disconnected. Nomophobia is currently considered a situational phobia [13]. In more details, nomophobia can be characterized as:

> *The fear of not being able to use a smartphone or a mobile phone and/or the services it offers...the fear of not being able to communicate, losing the connectedness that smartphones allow, not being able to access information through smartphones, and giving up the convenience that smartphones provide [2]*

Symptoms that characterize nomophobia include an excessive use of a mobile phone, which is kept permanently switched on, with the subsequent feeling of anxiety at the thought of a lack of network coverage. Other symptoms include the habit of continuously looking at the mobile screen in order to check for messages or missed calls (*ringxiety*, a combination of the words "ring" and "anxiety"), and the false sensation of hearing a mobile ring or vibration (the so-called "phantom vibration syndrome") [13]. To the best of our knowledge, no psychometric scales are available in the Italian language for investigating this psychological construct. We therefore planned a translation and validation study of the Nomophobia Questionnaire (NMP-Q), which is an instrument developed by Yildirim and Correia [2].

Methods

Instrument

The NMP-Q was translated from English into Italian using a classical "backwards and forwards" procedure. The questionnaire was then administered, along with a general questionnaire comprising basic sociodemographic information (age, gender, schooling level) and average smartphone use.

Study Design and Participants Selection

This investigation was designed as a cross-sectional study. Individuals were recruited via an online survey using a snowball approach that exploited Google Forms, which is an open-source tool for developing online questionnaires [14]. Given the exploratory nature of the study, a convenience sampling strategy was used. Online questionnaires were mainly circulated among undergraduate students and younger subjects, based on the idea that they were more likely to depend on their smartphones compared to other demographics. All procedures performed in the present study were in accordance with the ethical standards of the institutional research committee, and with the 1964 Helsinki declaration and its later amendments.

Statistical Analysis

Continuous data were represented as means and standard deviations (SDs), while categorical data were expressed as percentages. Skewness and kurtosis were computed for each item score. Acceptable values for asymmetry/skewness and kurtosis are in the range from -2 to +2 in case of normal univariate data distribution. To investigate the factor structure of the translated questionnaire, an exploratory factor analysis (EFA) was carried out, using the principal component analysis (PCA) approach with *varimax* rotation performed on the 20 items of the questionnaire. *Varimax* rotation was chosen since it minimizes factor complexity while, at the same time, maximizing the variance of factor loadings [15]. Different PCA runs were conducted. First, an exploratory PCA was performed on the 20 items of the questionnaire without carrying out any rotation, in order to (1) check whether PCA could be deemed an appropriate technique for the matrix by examining if the correlations among items were >.30, and (2) control for the factorability of the correlation matrix using Bartlett's test of sphericity. In cases of statistical significance, this test enables researchers to reject the null hypothesis (that is to say the correlations in the correlation matrix are zero and the matrix is an identity matrix).

The Kaiser-Meyer-Olkin (KMO) measure was calculated to assess the sampling adequacy. Ideally, the KMO should be >60.

The likely number of factors was determined by (1) the number of factors with eigenvalues greater than 1 [15,16], and (2) a visual inspection of the scree plot. After checking the factor loadings, items were deleted in cases of unsatisfactory loading (values less than .45) or loading conflicting with a sound theoretical explanation. Different PCAs with *varimax* rotation runs were, therefore, carried out iteratively until a satisfactory, clearly interpretable solution was finally achieved. Cases of cross-loading were interpreted according to salience and explained variance, with theoretical considerations also being taken into account. Internal consistency and reliability was computed by calculating the Cronbach alpha coefficient for the scale and each found subscale. To interpret the alpha coefficient, the following rule of thumb was used [17,18]: excellent if >.9, good in the range .8-.9, acceptable in the range .7-.8, questionable or adequate in the range .6-.7, poor in the range .5-.6 and unacceptable if <.5. Multivariate regression analyses were computed to shed light on the psychological predictors of nomophobia. All statistical analyses were performed using the IBM Statistical Package for the Social Sciences (SPSS, version 22, NY, USA). Figures with *P* values less than .05 were considered statistically significant.

Results

A total of 403 subjects (average age 27.91 years, SD 8.63; 160 males and 243 females, representing 39.7% and 60.3% of the sample, respectively) volunteered to take part in the study and were administered the Italian version of the NMP-Q (Textbox 1).

Regarding the schooling level, 2 subjects (2/403, 0.5%) had completed only elementary school, while 39 and 173 had completed middle school (39/403, 9.7%) and high school (173/403, 42.9%), respectively. All other subjects (189/403, 46.9%) had received higher education.

Concerning average smartphone use, 45 subjects usually spent less than 1 hour on their mobile phone (45/403, 11.2%), 94 spent between 1 and 2 hours (94/403, 23.3%), 69 spent between 2 and 3 hours (69/403, 17.1%), 58 spent between 3 and 4 hours (58/403, 14.4%), 48 spent between 4 and 5 hours (48/403, 11.9%), 29 spent between 5 and 7 hours (29/403, 7.2%), 36 spent between 7 and 9 hours (36/403, 8.9%) and 24 spent more than 10 hours (24/403, 6.0%). The means, SDs, and skewness and kurtosis figures are reported in Table 1. This information confirmed the normality of data distribution.

Since 403 subjects constituted a sample size large enough to compute reliable estimations of correlations among variables, we proceeded with the EFA. The N/k ratio (number of subjects/number of items) was 20.15:1, thus satisfying the requirements for commencing a factor analysis. Barlett's test of sphericity was significant (Chi-square=5,796.275, degrees of freedom=19, *P*<.001) and the KMO index was .941.

Textbox 1. Italian version of the Nomophobia Questionnaire by Yildirim and Correia.

1. Mi sento a disagio senza poter accedere costantemente alle informazioni tramite il mio smartphone
2. Sono infastidito/a se non riesco a cercare informazioni sul mio smartphone quando voglio farlo
3. Non essere in grado di ricevere le notizie (ad esempio gli ultimi aggiornamenti su eventi, meteo, ecc) sul mio smartphone mi rende nervoso/a
4. Sono seccato/a se non posso usare il mio smartphone e/o le sue applicazioni quando voglio farlo
5. L'idea di rimanere a corto di batteria nel mio smartphone mi spaventa
6. Se sono a corto di credito o se ho esaurito il mio limite di giga mensile, mi prende il panico
7. Se non c'è campo o non posso connettermi al Wi-Fi, rimango sempre a controllare per vedere se c'è segnale o se riesco a connettermi a una rete Wi-Fi
8. Se non posso usare il mio smartphone, ho paura di rimanere bloccato/a da qualche parte
9. Se non ho potuto controllare il mio smartphone per un po' di tempo, avverto il desiderio di controllarlo

Se non ho il mio smartphone con me,

10. Mi sento in ansia perché non riesco a comunicare istantaneamente con la mia famiglia e/o con gli amici
11. Sono preoccupato/a perché la mia famiglia e/o gli amici non possono raggiungermi
12. Mi sento nervoso/a perché non sono in grado di ricevere messaggi di testo e chiamate
13. Sono in ansia perché non riesco a rimanere in contatto con la mia famiglia e/o con gli amici
14. Sono nervoso/a perché non riesco a sapere se qualcuno mi ha cercato
15. Mi sento in ansia perché la mia connessione costante con la mia famiglia e gli amici è come se fosse rotta
16. Sono nervoso/a perché mi sento disconnesso/a dalla mia identità online
17. Sono a disagio perché non posso rimanere aggiornato/a con gli ultimi sviluppi dei social media e dei siti on-line
18. Mi sento a disagio perché non riesco a controllare le notifiche per gli aggiornamenti dei miei collegamenti e reti online
19. Mi sento in ansia perché non riesco a controllare i miei messaggi e-mail
20. Mi sento strano/a perché non saprei cosa fare

Table 1. Mean scores with standard deviation, skewness, and kurtosis.

Item Number	Mean	Standard deviation	Skewness	Kurtosis
1	3.610	1.720	0.122	–0.997
2	4.241	1.759	–0.263	–0.970
3	2.744	1.582	0.668	–0.426
4	3.945	1.775	–0.024	–1.019
5	3.663	1.790	0.179	–0.938
6	2.519	1.621	0.780	–0.484
7	2.732	1.694	0.938	0.063
8	2.390	1.631	1.126	0.428
9	3.975	1.793	–0.025	–0.972
10	3.591	1.835	0.132	–1.065
11	3.846	1.858	–0.024	–1.103
12	3.417	1.786	0.271	–1.016
13	3.588	1.825	0.172	–1.065
14	3.434	1.806	0.308	–0.944
15	2.675	1.716	0.762	–0.466
16	1.921	1.371	1.594	1.913
17	2.141	1.518	1.297	0.796
18	2.184	1.563	1.326	0.876
19	2.600	1.725	0.907	–0.189
20	2.285	1.609	1.158	0.406

As a result of the initial solution, four factors explaining up to 65.90% of the total variance were extracted. Before *varimax* rotation was performed, the first factor explained 23.32% of the variance, while the second and third explained 23.91% and 18.67%, respectively. After rotation, the three factors explained 23.32%, 23.91%, and 18.67% of the variance, respectively. The loadings of all items on each factor are shown in Table 2. Table 3 provides a summary of the EFA results and the reliability analysis of all items.

As seen in Table 2, the results of the second PCA run showed that each item loaded on a single factor, and that there was no cross-loading on other factors except for Item 5, which had a loading value of .452 on D1 and .520 on D3. This item was considered to load on its primary factor because this loading was more salient, thus explaining more variance, and was more theoretically reasonable. Furthermore, this value reached the cut-off value of .45; therefore, its loading on the secondary factor was not considered.

The Italian version of the NMP-Q showed an overall Cronbach alpha coefficient of .95 (.94, .89, and .88 for the three factors). As such, the internal consistency of the questionnaire can be considered good to excellent. The effect of dropping each variable per time is shown in Table 3. Removing each item per time leads to a decreased alpha coefficient for each item, showing that all items are essential to guarantee the questionnaire's reliability.

The NMP-Q total score correlated with the number of hours spent using a mobile phone (standardized beta-coefficient=.385, $P<.001$), and with gender (borderline statistical significance: standardized beta-coefficient=.091, $P=.057$), whereas no statistically significant associations could be detected with age and schooling level (Table 4). Similar patterns were found for the total score of the NMP-Q subscale "not being able to access information," with statistically significant associations with the number of hours spent on a smartphone (standardized beta-coefficient=.362, $P<.001$), gender (borderline statistical significance: standardized beta-coefficient=.088, $P=.067$) and schooling level (standardized beta-coefficient=.108, $P=.024$). The total score of the NMP-Q subscale "giving up convenience/losing connectedness" showed correlation only with the number of hours spent on a smartphone (standardized beta-coefficient=.402, $P<.001$). Finally, the score of the last NMP-Q subscale "not being able to communicate" correlated with the number of hours spent on a mobile device (standardized beta-coefficient=.254, $P<.001$) and with gender (standardized beta-coefficient=.142, $P=.004$).

In summary, the number of hours spent on a mobile phone turned out to be a predictor of all subscales, while gender was associated with D1 (although borderline significant) and D3. Finally, the schooling level correlated with D1. Further details are reported in Table 4.

Table 2. Factor loading of the Nomophobia Questionnaire. Salient factor loadings are indicated in italics.

Item Number	D1	D2	D3
1	0.255	0.307	*0.701*
2	0.235	0.110	*0.857*
3	0.201	0.301	*0.644*
4	0.257	0.186	*0.793*
5	0.452	0.292	*0.520*
6	0.273	*0.565*	0.368
7	0.143	*0.604*	0.359
8	0.366	*0.519*	0.105
9	0.334	0.394	*0.521*
10	*0.808*	0.255	0.244
11	*0.865*	0.078	0.255
12	*0.732*	0.299	0.356
13	*0.873*	0.159	0.272
14	*0.708*	0.369	0.328
15	*0.706*	0.438	0.137
16	0.265	*0.812*	0.108
17	0.163	*0.818*	0.198
18	0.156	*0.865*	0.194
19	0.235	*0.502*	0.253
20	0.136	*0.717*	0.208

Table 3. Cronbach alpha coefficient when one item is dropped.

Variable dropped	Raw alpha	Change in raw alpha	Standardized alpha	Change in standardized alpha
1	0.9420	-0.002916	0.9421	-0.002891
2	0.9428	-0.002099	0.9429	-0.002036
3	0.9431	-0.001823	0.9432	-0.001779
4	0.9424	-0.002526	0.9425	-0.002475
5	0.9417	-0.003142	0.9419	-0.003078
6	0.9422	-0.002689	0.9422	-0.002763
7	0.9432	-0.001728	0.9431	-0.001808
8	0.9440	-0.0009210	0.9440	-0.0009085
9	0.9419	-0.003009	0.9420	-0.002976
10	0.9409	-0.003969	0.9412	-0.003782
11	0.9422	-0.002717	0.9424	-0.002575
12	0.9402	-0.004725	0.9404	-0.004517
13	0.9410	-0.003918	0.9413	-0.003683
14	0.9399	-0.004983	0.9402	-0.004768
15	0.9412	-0.003729	0.9413	-0.003698
16	0.9423	-0.002553	0.9420	-0.002922
17	0.9423	-0.002538	0.9422	-0.002754
18	0.9420	-0.002886	0.9418	-0.003143
19	0.9443	-0.0006234	0.9443	-0.0006999
20	0.9434	-0.001453	0.9418	-0.003143

Table 4. Multivariate regression analysis investigating the impact of variables (age, gender, schooling level) on total and subscale scores.

Model	B	Standard deviation	Beta	*T*	*P* value
Overall					
(Constant)	31.819	7.413		4.292	.000
Number of hours spent using a mobile phone	4.450	0.590	.385	7.541	.000
Age	0.105	0.141	.038	0.747	.455
Gender	4.403	2.310	.091	1.906	.057
Schooling level	0.801	1.352	.028	0.593	.554
Not being able to access information					
(Constant)	10.722	2.556		4.195	.000
Number of hours spent using a mobile phone	1.435	0.203	.362	7.055	.000
Age	8.142E-005	0.049	.000	0.002	.999
Gender	1.463	0.796	.088	1.837	.067
Schooling level	1.052	0.466	.108	2.258	.024
Giving up convenience/losing connectedness					
(Constant)	11.518	2.955		3.897	.000
Number of hours spent using a mobile phone	1.849	0.235	.402	7.860	.000
Age	0.044	0.056	.040	0.781	.436
Gender	0.196	0.921	.010	0.213	.832
Schooling level	-0.375	0.539	-.033	-0.697	.486
Not being able to communicate					
(Constant)	9.579	3.056		3.135	.002
Number of hours spent using a mobile phone	1.165	0.243	.254	4.790	.000
Age	0.061	0.058	.056	1.057	.291
Gender	2.745	0.952	.142	2.882	.004
Schooling level	0.125	0.557	.011	0.223	.823

The impact on each item of age, gender, schooling level, and number of hours spent using a mobile phone is shown in Multimedia Appendix 1. Age had an effect on items 9 (borderline statistical significance: *P*=.063), 10 (borderline statistical significance: *P*=.089), and 15 (*P*=.003). The number of hours spent on one's own mobile device affected almost all items (1-4, 6-10, 15-20), whilst for items 5, 11, and 12 a borderline statistical significance could be found, and items 13 and 14 had no significance. Gender impacted only items 8 and 11 in a statistically borderline way (*P*=.087 and *P*=.075, respectively). Finally, schooling level affected items 1 (*P*=.098), 2 (*P*=.074), 5 (*P*=.086), 15 (*P*=.085), and 19 (*P*=.086), whereas this parameter had a significant impact on items 9 (*P*=.021), 10 (*P*=.004), and 13 (*P*=.037).

The interaction between the number of hours spent on a mobile device and gender significantly affected only item 10 (*P*=.016), whereas this parameter resulted in statistically borderline results for items 2 (*P*=.095), 7 (*P*=.080), and 19 (*P*=.074). The interaction between the number of hours spent on one's own smartphone and schooling level resulted in significance for items 9 (*P*=.003), 10 (*P*=.044), 11 (*P*=.019), and 18 (*P*=.010), and showed statistically borderline significance for items 2 (*P*=.073), 12 (*P*=.057), and 17 (*P*=.056). The interaction between schooling level and gender significantly impacted items 2-5. Finally, the interaction between the number of hours spent on a mobile phone, gender, and schooling level was statistically significant for items 3 (*P*=.011), 10 (*P*=.045), and 13 (*P*=.041), whereas they resulted in borderline significance for items 4 (*P*=.067), 16 (*P*=.066), and 20 (*P*=.053).

Discussion

Principal Findings

Based on the results of the reliability analysis, the internal consistency coefficient (Cronbach alpha) for the scale and all subscales of the NMP-Q was good, demonstrating that the NMP-Q is able to generate reliable scores. This finding is comparable with the alpha coefficient of the original instrument (Cronbach coefficient of .945, range of Cronbach coefficient for subscales from .819 to .939) [2]. The eigenvalues and scree plot supported a 3-factorial nature of the translated questionnaire.

These factors partially corresponded to the four factors found by Yildirim and Correia [2], namely items 1-5 and 9 ("not being able to access information"), items 6-8 ("giving up convenience"), items 16-20 ("losing connectedness"), and items

10-15 (“not being able to communicate”). The association between the scores and the number of hours spent using a mobile phone further corroborates and strengthens the validation of the NMP-Q in its Italian version. It is interesting to note that the gender effect using our instrument only had a statistically borderline influence. The few available studies and surveys that specifically assessed this point offer conflicting evidence [19,20]. This aspect warrants further investigations, dealing specifically with the gender issue.

The large sample size and the findings obtained in terms of psychometric properties represent the major strengths of our study. However, our research suffers from a number of limitations that should be properly recognized. First, since the sample is not representative, caution should be taken when making generalizations about our results. Further research should seek to replicate the results of the present study using more representative samples. Moreover, as for any other self-reported questionnaire, the self-reported structure of the NMP-Q may be a limitation due to social desirability bias. Nevertheless, there is a dearth of studies on nomophobia and this study makes it possible for researchers to use a reliable and validated instrument.

Conclusions

Nomophobia is a modern, emerging, situational, mobile phone-related phobia. The Italian version of the NMP-Q was validated and its psychometric properties were examined, showing a 3-factor structure. The Italian NMP-Q proved to be reliable and can therefore be employed by researchers. Further studies are needed to assess the consistency of the NMP-Q in other samples (either general or clinical), and to investigate comorbidities and predictors of nomophobia using a confirmatory factorial approach to obtain more robust results. The relationship of nomophobia with other ICT-related psychological disorders (such as the Internet addiction) also warrants further investigations.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Impact of age, gender, number of hours spent using a mobile phone, and schooling level on items scores.

[PDF File (Adobe PDF File), 28KB - mhealth_v6i1e24_app1.pdf]

References

1. Liu CH, Lin SH, Pan YC, Lin YH. Smartphone gaming and frequent use pattern associated with smartphone addiction. Medicine (Baltimore) 2016 Jul;95(28):e4068 [FREE Full text] [doi: 10.1097/MD.0000000000004068] [Medline: 27428191]
2. Yildirim C, Correia AP. Exploring the dimensions of nomophobia: development and validation of a self-reported questionnaire. Comput Human Behav 2015;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
3. Pew Research Center. 2015 Apr 01. U.S. smartphone use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2018-01-17] [WebCite Cache ID 6wYBnjMhA]
4. iStat. 2014 Dec 18. Cittadini e nuove tecnologie URL: http://www.istat.it/it/files/2014/12/Cittadini_e_nuove_tecnologie_anno-2014.pdf [accessed 2018-01-17] [WebCite Cache ID 6wYBx0F7f]
5. Kesari KK, Siddiqui MH, Meena R, Verma HN, Kumar S. Cell phone radiation exposure on brain and associated biological systems. Indian J Exp Biol 2013 Mar;51(3):187-200. [Medline: 23678539]
6. Makker K, Varghese A, Desai NR, Mouradi R, Agarwal A. Cell phones: modern man's nemesis? Reprod Biomed Online 2009 Jan;18(1):148-157. [Medline: 19146782]
7. Flaherty GT, Choi J. The 'selfie' phenomenon: reducing the risk of harm while using smartphones during international travel. J Travel Med 2016 Feb;23(2):tav026. [doi: 10.1093/jtm/tav026] [Medline: 26858274]
8. Moussa MM. Review on health effects related to mobile phones. Part II: results and conclusions. J Egypt Public Health Assoc 2011;86(5-6):79-89. [doi: 10.1097/01.EPX.0000406204.36949.49] [Medline: 22173110]
9. Neuenschwander M. Online addictive disease. Ther Umsch 2014 Oct;71(10):599-607. [doi: 10.1024/0040-5930/a000559] [Medline: 25257114]
10. Ascher MS. In: Levounis P, editor. The Behavioral Addictions. Washington, DC: American Psychiatric Association Publishing; 2015.
11. Wright PJ, Donnerstein E. Sex online: pornography, sexual solicitation, and sexting. Adolesc Med State Art Rev 2014 Dec;25(3):574-589. [Medline: 27120886]
12. Karadağ E, Tosuntaş Ş, Erzen E, Duru P, Bostan N, Şahin BM, et al. Determinants of phubbing, which is the sum of many virtual addictions: a structural equation model. J Behav Addict 2015 Jun;4(2):60-74 [FREE Full text] [doi: 10.1556/2006.4.2015.005] [Medline: 26014669]
13. Bragazzi NL, Del PG. A proposal for including nomophobia in the new DSM-V. Psychol Res Behav Manag 2014 May 16;7:155-160 [FREE Full text] [doi: 10.2147/PRBM.S41386] [Medline: 24876797]
14. Google. Google Forms. 2018. URL: https://www.google.it/intl/it/forms/about/ [accessed 2018-01-17] [WebCite Cache ID 6wYCP9fUG]

15. Tabachnick BG, Fidell LS. Using multivariate statistics. Boston, MA: Pearson; 2013.
16. Field A. Discovering statistics using SPSS. Beverly Hills, CA: Sage; 2009.
17. Nunnally JC. Psychometric theory, 2nd edition. New York: McGraw-Hill; 1978.
18. George D, Mallery P. SPSS for Windows Step by Step: A Simple Guide and Reference, 18.0 update. Boston, MA: Allyn & Bacon/Pearson; 2011.
19. Daily Mail Online. 2008 Mar 31. Nomophobia is the fear of being out of mobile phone contact - and it's the plague of our 24/7 age URL: http://www.dailymail.co.uk/news/article-550610/Nomophobia-fear-mobile-phone-contact--plague-24-7-age.html [accessed 2018-01-10] [WebCite Cache ID 6wMhU6EWt]
20. SecurEnvoy. 2012 Feb 16. 66% of the population suffer from nomophobia the fear of being without their phone URL: https://www.securenvoy.com/blog/2012/02/16/66-of-the-population-suffer-from-nomophobia-the-fear-of-being-without-their-phone/ [accessed 2018-01-10] [WebCite Cache ID 6wMiOyD1g]

Abbreviations

EFA: exploratory factor analysis
ICT: information and communication technology
KMO: Kaiser-Meyer-Olkin measure
NMP-Q: Nomophobia Questionnaire
PCA: principal component analysis

Edited by G Eysenbach; submitted 15.10.17; peer-reviewed by J Jones, K Agarwal, P Boisrond; comments to author 17.11.17; revised version received 11.12.17; accepted 02.01.18; published 22.01.18.

Please cite as:
Adawi M, Bragazzi NL, Argumosa-Villar L, Boada-Grau J, Vigil-Colet A, Yildirim C, Del Puente G, Watad A
Translation and Validation of the Nomophobia Questionnaire in the Italian Language: Exploratory Factor Analysis
JMIR Mhealth Uhealth 2018;6(1):e24
URL: http://mhealth.jmir.org/2018/1/e24/
doi:10.2196/mhealth.9186
PMID:29358161

Original Paper

Development and Evaluation of a Mobile Personalized Blood Glucose Prediction System for Patients With Gestational Diabetes Mellitus

Evgenii Pustozerov[1,2*], MSc; Polina Popova[2,3*], MD, PhD; Aleksandra Tkachuk[2], MD; Yana Bolotko[2], MD; Zafar Yuldashev[1], DSc, PhD; Elena Grineva[2,3], DSc, PhD, MD

[1]Department of Biomedical Engineering, Saint Petersburg State Electrotechnical University, Saint Petersburg, Russian Federation

[2]Institute of Endocrinology, Almazov National Medical Research Centre, Saint Petersburg, Russian Federation

[3]Department of Internal Diseases and Endocrinology, Pavlov First Saint Petersburg State Medical University, Saint Petersburg, Russian Federation

[*]these authors contributed equally

Corresponding Author:
Evgenii Pustozerov, MSc
Department of Biomedical Engineering
Saint Petersburg State Electrotechnical University
Professora Popova, 5
Saint Petersburg, 197376
Russian Federation
Phone: 7 812 234 01 33
Fax: 7 812 346 27 5
Email: pustozerov.e@gmail.com

Abstract

Background: Personalized blood glucose (BG) prediction for diabetes patients is an important goal that is pursued by many researchers worldwide. Despite many proposals, only a few projects are dedicated to the development of complete recommender system infrastructures that incorporate BG prediction algorithms for diabetes patients. The development and implementation of such a system aided by mobile technology is of particular interest to patients with gestational diabetes mellitus (GDM), especially considering the significant importance of quickly achieving adequate BG control for these patients in a short period (ie, during pregnancy) and a typically higher acceptance rate for mobile health (mHealth) solutions for short- to midterm usage.

Objective: This study was conducted with the objective of developing infrastructure comprising data processing algorithms, BG prediction models, and an appropriate mobile app for patients' electronic record management to guide BG prediction-based personalized recommendations for patients with GDM.

Methods: A mobile app for electronic diary management was developed along with data exchange and continuous BG signal processing software. Both components were coupled to obtain the necessary data for use in the personalized BG prediction system. Necessary data on meals, BG measurements, and other events were collected via the implemented mobile app and continuous glucose monitoring (CGM) system processing software. These data were used to tune and evaluate the BG prediction model, which included an algorithm for dynamic coefficients tuning. In the clinical study, 62 participants (GDM: n=49; control: n=13) took part in a 1-week monitoring trial during which they used the mobile app to track their meals and self-measurements of BG and CGM system for continuous BG monitoring. The data on 909 food intakes and corresponding postprandial BG curves as well as the set of patients' characteristics (eg, glycated hemoglobin, body mass index [BMI], age, and lifestyle parameters) were selected as inputs for the BG prediction models.

Results: The prediction results by the models for BG levels 1 hour after food intake were root mean square error=0.87 mmol/L, mean absolute error=0.69 mmol/L, and mean absolute percentage error=12.8%, which correspond to an adequate prediction accuracy for BG control decisions.

Conclusions: The mobile app for the collection and processing of relevant data, appropriate software for CGM system signals processing, and BG prediction models were developed for a recommender system. The developed system may help improve BG control in patients with GDM; this will be the subject of evaluation in a subsequent study.

(JMIR Mhealth Uhealth 2018;6(1):e6) doi:10.2196/mhealth.9236

KEYWORDS

blood glucose prediction; mHealth; gestational diabetes mellitus; recommender system; personalized medicine; mobile app

Introduction

Gestational diabetes mellitus (GDM) is one of most common endocrine disorders during gestation, affecting up to 17.8% of pregnancies [1]. It is associated with short-term obstetric and perinatal complications such as preeclampsia, increased cesarean delivery rates, macrosomia, and birth injury [1], as well as long-term future metabolic health implications for the mother and the offspring (ie, increased risk of obesity and type 2 diabetes) [2]. Thus, maintaining normal blood glucose (BG) levels during pregnancy is critical for preventing adverse pregnancy outcomes and to stop the cycle that perpetuates the transmission of metabolic disease to the offspring [3].

The timeframe for effective interventions to prevent complications from GDM is usually limited to the third trimester of pregnancy and the physiology of pregnancy is rapidly changing (eg, increasing insulin resistance); therefore, women with GDM require frequent visits to health care providers to ensure good glycemic control (usually every 2-4 weeks on diet and every 1-2 weeks when treated with insulin). These frequent antenatal visits are both a considerable burden to the patients and, considering the increasing incidence of GDM, may also place significant stress on health care systems and their often-limited resources.

The technology used to remotely deliver health care—specifically, via an electronic means of communication (eg, mobile health [mHealth])—offers an appealing solution to this problem. Currently, an increasing number of articles are reporting on mobile apps for patients with different types of diabetes. Several involve mobile phone-based randomized controlled trials (RCTs) that show promising results for the self-management of diabetes [4-6].

To the best of our knowledge, no completed RCTs that assessed the effectiveness of mobile apps in GDM patients have been published, although there are several ongoing trials [7]. Approximately 70% to 85% of GDM patients can control GDM with lifestyle modification alone [8]; consequently, an effective tool for making appropriate food choices to prevent high BG levels as postprandial glucose (PPG) responses would be of particular importance for women with GDM.

The current methods for predicting PPG responses to food, albeit important in the context, are limited and imprecise. Basing predictions on meal carbohydrate content is the most common method [9], but it is not sufficiently precise in predicting PPG response [10].

Other methods of estimating PPG responses are glycemic index and glycemic load [11]. However, because both methods are based on the assessment of PPG response to the consumption of certain kinds of food, it is difficult to apply them in clinical practice when there are different food combinations and proportions [12]. Furthermore, reliable databases describing the glycemic index of different foods are absent in many countries. Moreover, several studies have found high variability in individuals' glycemic responses to meals with identical nutritional composition [10,13], but the reasons for this variability are not yet sufficiently clear. PPG responses may depend on individual lifestyle [14], genetics [15], glucose transporter activity levels [16], and gut microbiota [10].

Zeevi et al [10] developed a machine learning algorithm that integrates multidimensional data on blood parameters, anthropometrics, physical activity, self-reported lifestyle behaviors, and gut microbiota composition to predict personalized PPG responses in healthy individuals. However, adoption of their algorithm in clinical practice may be limited by its complexity and the high cost of microbiota analyses. Further, it has not been validated on pregnant women and its usability for prediction of PPG responses during pregnancy is not known.

To the best of our knowledge, little work has been conducted on the development of PPG response prediction models for GDM patients, as the primary task for researchers remains prediction of BG levels for patients with type 1 diabetes. Further, although numerous BG prediction algorithms have been proposed, only a few projects are dedicated to the development of complete recommender system infrastructures that incorporate BG prediction algorithms for diabetes patients [17]. The development and implementation of such a system might be particularly important for patients with GDM, especially considering the high importance of BG control for these patients during pregnancy and a typically higher acceptance rate for mHealth solutions for short- to midterm usage.

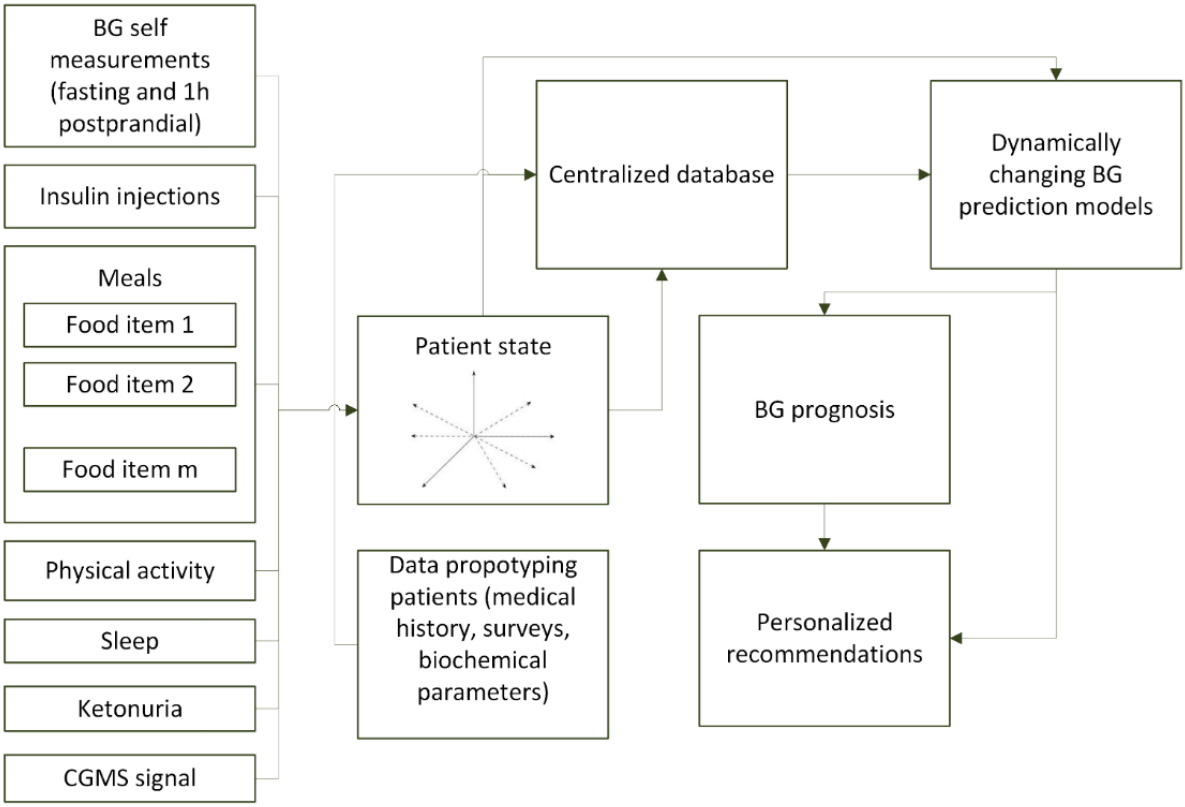
Therefore, the goal of this study was to develop an infrastructure that incorporates data processing algorithms, BG prediction models, and an appropriate mobile app for patients' electronic record management to guide BG prediction-based personalized recommendations for GDM patients.

Methods

Analysis

Following analysis of the literature on the available apps for diabetes management [18], main design functions for diabetes apps [19,20], app evaluations [21], and user reviews of the top diabetes apps available in the Google Play and Apple Store markets, we defined the most important points to consider within our project. For a mobile monitoring system, we derived the following core features: (1) systematic collection and persistence of patient data locally on the user device and centrally on a server, (2) data export into commonly used formats for assessment by patient and doctor, and (3) personalized recommendations based on evaluation of PPG response. Further, for the system to be sufficient for long-term monitoring, the following requirements must be satisfied: the system must be based on a commonly used architecture (in terms of devices, software, etc), the system must be sufficiently simple to use by patients without prior computer literacy, and the process of data collection and exchange should be as simple as possible.

Figure 1. Conceptual scheme of the gestational diabetes mellitus recommender system. BG: blood glucose; CGM: continuous glucose monitoring.



The conceptual scheme of the GDM recommender system is shown in Figure 1. The core of the recommender system comprises dynamically changing BG predicting models, which are used for personalized recommendations. The architecture utilizes the data collected from different sources for all patients to constantly improve its BG predictions. The patient state is a vector in multidimensional space and contains data on preceding events (eg, meals and insulin injections) and information on BG levels collected within the continuous glucose monitoring (CGM) system signal. The patient state is combined with data prototyping patients (with the data relevant to BG regulation, such as glycated hemoglobin A_{1c} [HbA_{1c}] and oral glucose tolerance test [OGTT] parameters) and the data for all patients in a centralized database that is used to train the BG prediction models. The models are dynamically retrained as new data are uploaded making their predictions more sustainable when they are used by new patients for whom only a limited amount of data has so far been recorded.

Design

In the first step of the system design, we formulated a list of parameters for the BG predicting algorithm (Textbox 1). Because it was not feasible to track these records in a traditional paper diary and there were no solutions matching our requirements, such as apps for recording meals in a simple and hassle-free (mobile) manner and corresponding software to acquire the records of these electronic diaries alongside CGM system signals, we developed our own app to obtain the necessary records for the BG prediction algorithm.

From a technical viewpoint, the developed GDM advisory system contained the following elements: (1) a mobile app for data collection and presentation on the side of the patient, (2) a CGM system for continuous BG monitoring, and (3) a centralized server with appropriate software for data aggregation, processing, and training of BG prediction models.

Textbox 1. Groups and subgroups of parameters necessary for the BG prediction algorithm.

Biomedical signals (implemented by devices)

- Continuous glucose monitoring system signal features describing postprandial glucose response

Electronic diaries records (implemented by mobile app)

- Blood glucose measurements, meals, insulin injections, physical activity, sleep duration

Individual patient characteristics

- Biometric characteristics, medical history and survey data, biochemical parameters

Development

Mobile and Desktop Diary App

The mobile app was developed using the Java programming language and it supports devices running Android OS 2.3 and higher. The desktop app was also developed using the Java programming language. It supports devices running Java 7.0 and higher. The app includes a food database created based on reference books of the Russian Academy of Medical Sciences and the US Department of Agriculture (USDA) Food Composition Databases (Release 28) [22]. Complex dishes included into the database were supplemented by recipes from reference books of the Russian Academy of Medical Sciences [23].

The development process comprised several iterations and app usability was evaluated with the help of 36 GDM patients. The patients were asked to provide feedback on app usability, evaluate its core features, indicate limitations they encountered, and possible ways to improve the app. Seven major corrections were made before the study, including improving the ergonomic aspects of the app, increasing the size of the built-in food database, and enhancement of app functionality, including respective functions for modifying previously recorded data, combining existing food items in the built-in food database into new recipes, and marking particular food items as favorites. All the preceding corrections were made before commencing the study.

Data Processing Software

The centralized software component was based on programs that operate in an automatic sequence, evoked by a script. The complete data processing algorithm, which transforms the raw data of the electronic reports from the patients and the CGM system signals of the iPro device, comprises the following five steps.

Step 1

The initial signal data was recorded by the CGM system (Medtronic iPro was used in this study) and uploaded to the CareLink server used by the Medtronic iPro system, where the signals for patients were stored. For the purpose of this study, the data of the participants were retrieved from the Medtronic server in the form of CSV files by our software and processed by an automatic script.

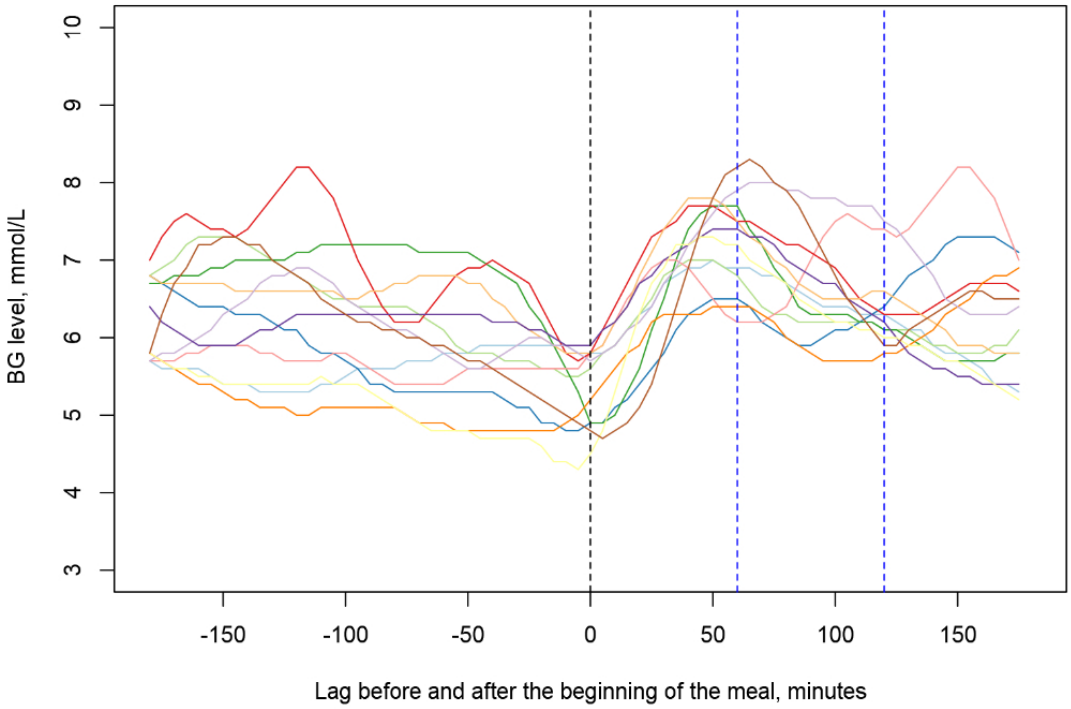
Step 2

Electronic diaries, exported from the mobile app and collected on the centralized server, were processed in such a manner that every meal was stored together with the parameters characterizing previous events (eg, any meal 3 hours before the current meal and any physical exercise or sleep).

Step 3

Records from the electronic diaries were matched with appropriate CGM system signals for each patient: for every food intake in a diary, 3 hours of BG signals before and after the meal were collected. The information on the BG curve was stored together with food records collected in step 2. Figure 2 shows an ensemble of resulting BG curves 3 hours before and after meals, collected over the span of a week.

Figure 2. Ensemble of blood glucose (BG) curves collected 3 hours before and after meals for one of the patients. Different colors represent different meals.



Step 4

Based on the postprandial BG curve, a set of parameters characterizing PPG response were calculated:

1. AUC60: area under the glycemic curve 1 hour after the start of the meal ([mmol/L]/hour);
2. AUC120: area under the glycemic curve 2 hours after the start of the meal ([mmol/L]/hour);
3. BG60: blood glucose level 1 hour after the start of the meal (mmol/L); and
4. Peak BG: peak value on a postprandial BG curve (mmol/L).

Step 5

The resulting data were supplemented by the data used to prototype the patients (medical history data, survey data, biochemical parameters). The data from different patients were combined in an integral data frame, which was the prepared input data for the BG prediction models.

Recruitment and Enrollment

This study was part of the ongoing clinical trial "Genetic and Epigenetic Mechanisms of Developing Gestational Diabetes Mellitus and its Effects on the Fetus" (GEM GDM; trial registration number: A A A A - A 16-116012210374-0), which started in July 2015. Participation in this study using the CGM system was optional for the participants of the GEM GDM trial. Pregnant women were invited to take part in this study if they were using our mobile app or our desktop app and provided accurate information concerning their food intake and BG measurements. Those who had pregestational diabetes and other diseases affecting carbohydrate metabolism were excluded. None of the participants were treated with insulin prior to or during this study. The study was approved by the ethical committee of the National Almazov Medical Research Centre, Saint-Petersburg, Russia (protocol no 119), and the participants gave their consent in writing.

In total, 66 of 158 women agreed to participate. Four women were excluded from the analysis because they provided inaccurate information on food intake during the week they wore the CGM system (see "Control of the Accuracy of Self-Reports"). As a result, 62 participants (48 pregnant women with GDM and 14 women with normal glucose tolerance) were included in the study.

The diagnosis of GDM was based on the Russian National Consensus [24] and the recommendations of the International Association of Diabetes and Pregnancy Study Groups (IADPSG) [25] based on the results of 2 hour OGTT performed during the 24th to 28th week of gestation. Pregnant women without diabetes were included as controls.

Measures

Glucose was measured for 7 days using the iPro2 CGM with Enlite sensors (Medtronic, Minneapolis, MN, USA) and independently calibrated with the Accu-Check Performa Nano blood glucose meter (Roche Diabetes Care, Indianapolis, IN, USA) for a minimum of four measurements per day.

During that week, participants were instructed to record all daily activities, including meals with exact components and weights, using our mobile app or our desktop app.

Prior to the study, the women were questioned about their clinical characteristics and completed a special questionnaire under supervision. The questionnaire consisted of the following sections: frequency of consumption of basic products in a week, physical activity, and smoking before and during pregnancy. The sections of the form were defined in a semiquantitative manner, reflecting different frequency levels of consuming certain products and performing physical activity (low, medium, and high). The description of these semiquantitative variables is presented in Multimedia Appendix 1. This questionnaire was previously reported [26].

Control of the Accuracy of Self-Reports

To avoid biases resulting from inaccurate self-reports about daily activities, especially meals, we took several precautions. The women were provided with kitchen scales to measure the weight of each kind of food consumed at home (in grams) and were asked to check the weight of meal components consumed in restaurants and other public catering places. The reports resulting from the collection of data on food intake were analyzed by endocrinologists and discussed with the participants in detail. If the BG curves collected via the iPro2 CGM showed that two or more food intakes per week were not documented in the app or that two or more BG measurements checked in the glucometer memory differed from the BG levels reported by a participant, the data were excluded from the analysis (four women).

Statistical Analysis

Data were statistically processed with SPSS 22.0 [27], MATLAB 2016r [28], R 3.4.0 [29], and Python 2.7.14 [30]. The data are presented as the mean and standard deviation. Differences in the quantitative characteristics of the groups were assessed with Student *t* test. The chi-square criterion was used to compare the distribution of qualitative characteristics. The differences were considered significant at $P<.05$.

Blood Glucose Prediction Model

After the data processing phase, the data were used to create the BG prediction model. We developed linear regression models with the use of lasso regularization for feature selection and coefficient tuning [31] to avoid overfitting. Linear regression was chosen because of its good interpretability, simplicity, rapid tuning, and adequate accuracy in comparison with other methods performing the task.

Participants were separated into training and test sets in the proportion of 80%:20% and 20-fold cross-validation was used for parameter tuning.

The optimization task for lasso regularization was $\| y-X\omega \|^2+\lambda \| \omega \| \rightarrow \min_{\omega}$, $\lambda \geq 0$, where y is a vector of output values, ω is a vector of weights, $X=[x_1,...,x_n]$ is a set of input values for all objects in the dataset, and λ is the tunable regularization coefficient (double vertical bars stand for a norm of a vector/matrix).

As a dependent indicator, the features of PPG response (AUC60, AUC120, BG60, and peak BG) were determined for appropriate models. The following parameters were imputed in the dataset as potential predictors of the features of PPG response:

1. Anthropometric parameters (eg, age, weight, body mass index [BMI], gestational age, and systolic and diastolic blood pressure),
2. Medical history data (impaired glucose tolerance; polycystic ovary syndrome; family history of diabetes; number of pregnancies, abortions, deliveries, and miscarriages; arterial hypertension; the use of combined oral contraceptive pills before pregnancy; and GDM in history), biochemical parameters (fasting, 1-hour and 2-hour BG levels at OGTT, fasting insulin, HbA_{1c}, total cholesterol level, very low density and high density lipoproteins, and triglycerides at the time of OGTT).
3. Survey data: 11 parameters associated with the consumption of certain product groups (fruits, pastries, skimmed dairy products, legumes, meat, sausage products, dried fruits, fish, whole grain bread, sauces, and vegetables), three parameters related to beverages (alcohol, sweet drinks, and coffee), and three parameters characterizing physical activity (walking, climbing the stairs, and performing sports). For each listed parameter, the intensity was estimated on an ordinal scale of three levels: low, medium, and high. Smoking was marked as "yes" or "no." Smoking, alcohol intake, and physical activity parameters were assessed separately before and during pregnancy. Because none of the participants reported the highest category of activity for frequency of sports activities (>3 days/week), frequency of climbing the stairs (>16 flights/day), and legume consumption (>2 portions/week) at the time of glucose self-monitoring, we recorded the remaining two categories into binary variants ("yes": medium category; "no": the lowest category of activity) for statistical analyses. Walking duration and coffee consumption were coded as variables with three levels: (0 for low, 1 for medium, and 2 for high).
4. Current and preceding meal: 33 parameters including type of food intake (1=breakfast, 2=lunch, 3=dinner, 4=snack); macronutrient and micronutrient content (water [g], energetic value [kcal], fats [g], carbohydrates [g], dietary fibers [g], sugars [g], calcium [mg], iron [mg], phosphorus [mg], zinc [mg], copper [mg], vitamin C [mg], riboflavin [mg], niacin [mg], thiamin [mg], vitamin B6 [mg], folate [mcg], folic acid [mcg], retinol [mcg], retinol equivalent [mcg], alpha-carotene [mcg], beta-carotene [mcg], vitamin E [mg], vitamin D [mg]); the presence of a preceding meal within 3 hours before the index meal (yes/no); and the amount and percentage of carbohydrates in the preceding meal.

The coefficients were tuned via a coordinate descent optimization algorithm (using the glmnet package for R) [32]. The λ coefficient for each model was chosen in such a manner as to obtain the smallest number of nonzero coefficients, at which the mean squared error, estimated during cross-validation, was in the range of one standard deviation from the best model fit. This allowed us to obtain a simple, yet sufficiently accurate model.

Results

Mobile App

The mobile app for data collection and exchange was developed for the Android OS. A desktop app with the same functionality was also developed for users not in possession of an Android device (both are referred to simply as "app").

The app contains 18 different screens, including the main menu, user input forms, record management and information, user settings, data export form, and help. Some of these screens are presented in the animation included in Multimedia Appendix 2.

The SQLite database in the app consisted of 13 tables (tables for records on BG, insulin, physical activity, sleep, ketones, meals, and meal items; tables with built-in and user food databases and user data) containing the data for different types of user records as well as the built-in food database. The built-in food database, collected from open sources (including the Scientific Research Institute of Nutrition of the Russian Academy of Medical Sciences and the USDA food databases), made it possible to track 27 food parameters (macronutrients and micronutrients) without patient input, because involving the patients could lead to mistakes and additional burden from the patients' perspective.

The app allowed the users to export data in Excel spreadsheets and store them on their devices as well as to send them remotely to physicians (Figure 3).

The Data Processing Algorithm

Using the preceding methods, an algorithm transformed the data for the amount and kind of consumed food, the start time of food intake, physical activity, duration of sleep, and current BG level (received from the CGM system) into a BG prediction parameter used to establish a recommender system. An example of the results of data matching between the CGM system signal and the electronic diary is shown in Figure 4 and more examples are shown in Multimedia Appendix 3.

Participants' Characteristics

The main characteristics of the participants are presented in Table 1. The women with GDM had higher BMI and higher levels of HbA_{1c} and plasma glucose (PG) during OGTT than the controls.

Figure 3. Example of a standardized report exported from the app.

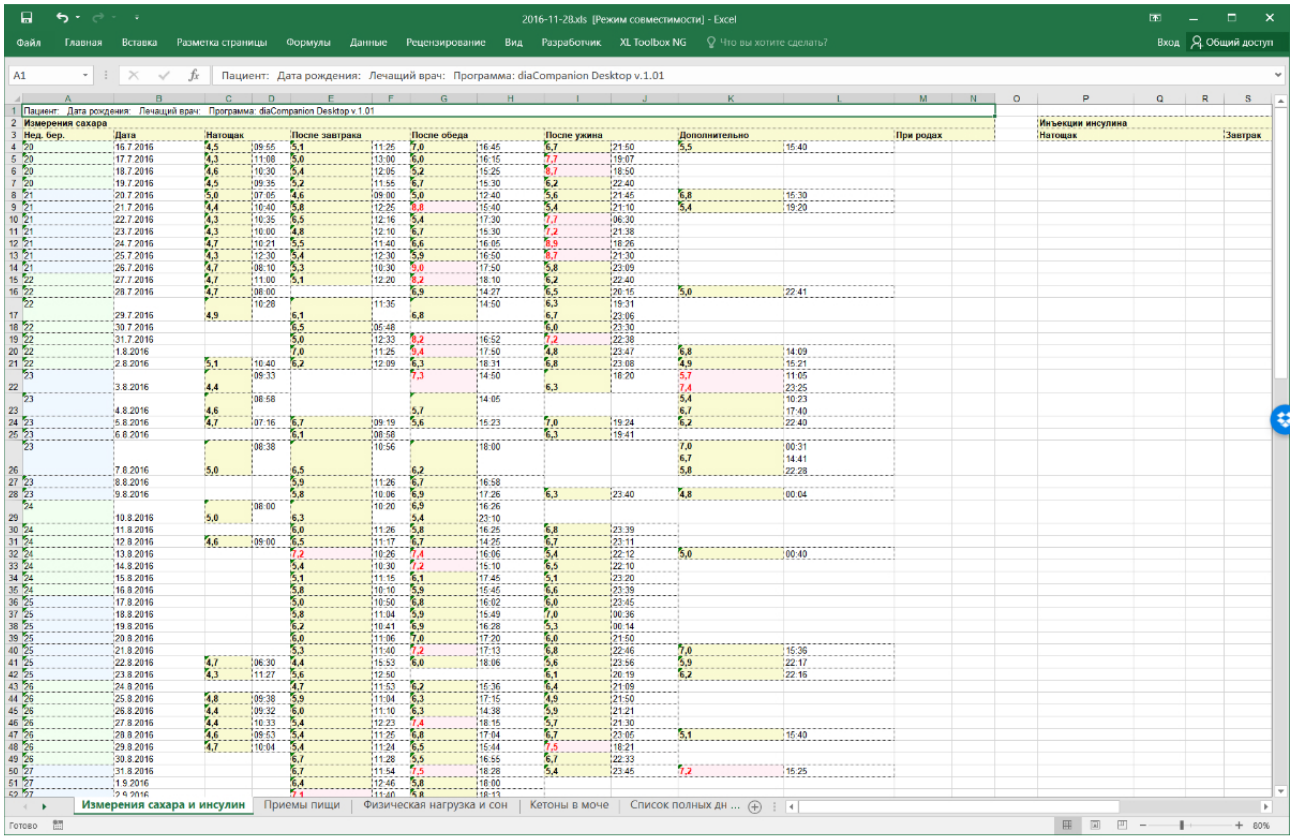


Figure 4. Result of data matching between the continuous glucose monitoring system signal and the electronic diary.

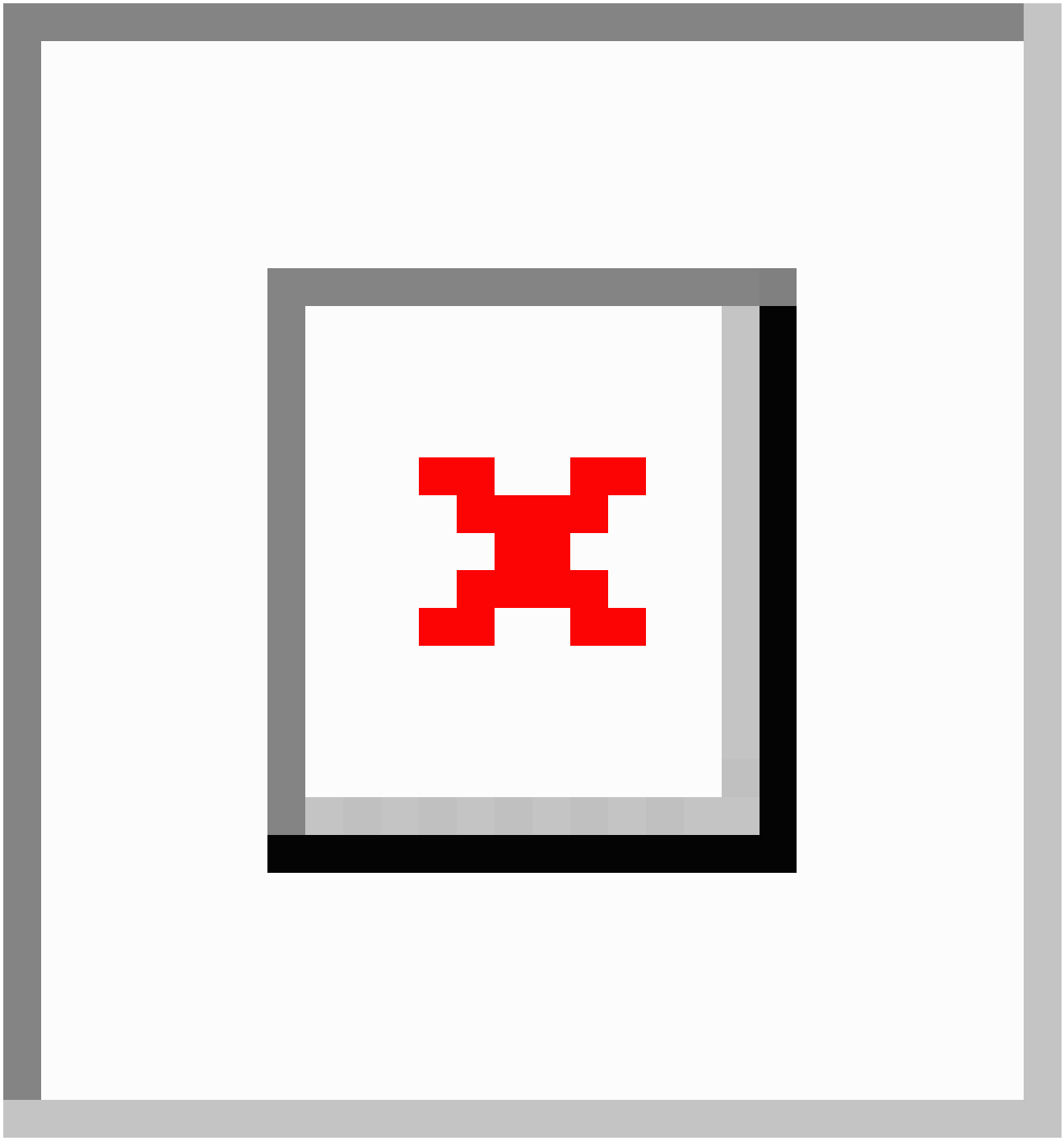


Table 1. Characteristics of the participants (N=62).

Characteristic	GDM (n=48)	Control (n=14)	*P* (two-sided test)
Age (years), mean (SD[a])	32.1 (4.0)	29.8 (2.9)	.06
Prepregnancy BMI[b] (kg/m^2), mean (SD)	26.4 (6.4)	21.1 (3.4)	.006
HbA_{1c}[c] (%), mean (SD)	5.13 (0.40)	4.84 (0.40)	.03
Gestational age (weeks), mean (SD)	31.4 (3.0)	31.4 (2.8)	>.99
BP[d] systolic (mm Hg), mean (SD)	118.9 (10.6)	117.5 (13.9)	.70
BP diastolic (mm Hg), mean (SD)	75.1 (7.9)	74.8 (14.9)	.90
Arterial hypertension, n (%)	20 (42)	5 (36)	.77
OGTT[e] fasting PG[f] (mmol/L), mean (SD)	5.0 (0.7)	4.2 (0.5)	<.001
OGTT 1-h PG (mmol/L), mean (SD)	9.6 (2.3)	6.3 (1.6)	<.001
OGTT 2-h PG (mmol/L), mean (SD)	8.4 (2.4)	5.5 (1.4)	<.001
Fasting serum insulin (pmol/L), mean (SD)	96.4 (52.8)	93.9 (81.9)	.91

[a]SD: standard deviation.

[b]BMI: body mass index.

[c]HbA_{1c}: glycated hemoglobin A_{1c}.

[d]BP: blood pressure.

[e]OGTT: oral glucose tolerance test.

[f]PG: plasma glucose.

Table 2. Glycemic response and meal characteristics for gestational diabetes mellitus (GDM) and control patients.

Characteristic	GDM, mean (SD[a])	Control, mean (SD)	*P* (two-sided test)
Fasting BG[b] (mmol/L)	5.1 (0.7)	5.0 (0.6)	<.001
BG60[c] (mmol/L)	6.2 (1.0)	5.9 (0.9)	<.001
AUC60[d] (mmol/L*hour)	5.77 (0.80)	5.62 (0.74)	.02
AUC120[e] (mmol/L*hour)	5.86 (0.78)	5.68 (0.70)	<.001
BG Rise 1h after meal (mmol/L)	1.5 (1.0)	1.6 (1.0)	.66
Postprandial peak BG (mmol/L)	6.6 (1.0)	6.5 (1.0)	.02
Time to peak BG (minutes)	75.0 (43.7)	73.6 (46.0)	.68
Carbohydrates per meal (g)	31.8 (22.2)	51.5 (31.5)	<.001
Proteins per meal (g)	22.5 (15.1)	22.9 (15.7)	.72
Fats per meal (g)	19.6 (15.1)	25.2 (17.0)	<.001
Energy per meal (kcal)	398 (209)	530 (279)	<.001

[a]SD: standard deviation.

[b]BP: blood pressure.

[c]BG60: blood glucose level 60 minutes after the meal.

[d]AUC60: area under the postprandial blood glucose curve 60 minutes after the meal

[e]AUC120: area under the postprandial blood glucose curve 120 minutes after the meal.

Characteristics of Meals and Glycemic Responses

The average characteristics of meals and glycemic responses are presented in Table 2. Patients with GDM consumed significantly lower amounts of carbohydrates and fats in their meals; therefore, the energy content of their meals was considerably lower than that of the control group. Fasting and postprandial BG levels were significantly higher in patients with GDM than in those in the control group, whereas the actual rise in BG level after meals did not vary significantly in these groups owing to lower average carbohydrate consumption in the GDM group. The area under the curve (AUC) for BG level 1 and 2 hours after the beginning of the meal was also larger for patients in the GDM group, even considering the lower average carbohydrate consumption in this group.

Blood Glucose Prediction Models

The significant model coefficients are presented in Table 3. The BG level 2 hours after OGTT was a heightening factor for all the variables describing glycemic response, corresponding to worse BG regulation in participants with GDM. Some of the parameters describing lifestyle during pregnancy were significant in predicting PPG response: reported physical activity was a lowering factor for all the variables describing glycemic response, where they were found to be significant, as well as reported high consumption of legumes, although high consumption of coffee appeared to be a heightening factor.

Despite their simplicity, the developed linear regression models proved to be highly efficient in their prediction of the PPG response feature. Model performance was estimated using standardized metrics. The correlation between real and predicted values (*R*), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) were estimated for each of the proposed models (Table 4). The resulting table shows only marginally worse results for the test set, which may be due to overfitting. When comparing the MAPE of the presented models, it might be considered that the models predicting AUC on a postprandial curve perform better.

Table 3. Coefficients of the linear regression models predicting different features of postprandial glucose (PPG) response.

Parameter	AUC60[a]	AUC120[b]	BG60[c]	Peak BG[d]
Intercept	1.6246	2.5650	2.1860	3.4590
1. BG level before meal (mmol/L)	0.6877	0.6033	0.4116	0.5959
2. Breakfast (yes/no)	0.2927	0.2337	0.2746	0.2832
3. Carbohydrates (g)	0.0030	0.0034	0.0072	0.0093
4. Starch (g)	—	0.0017	—	0.0024
5. Carbohydrates (%)	0.1951	0.0289	0.0902	—
6. Proteins (%)	—	—	—	–0.4503
7. Preceding meal (yes/no)	—	–0.0539	–0.1570	–0.0730
8. Carbohydrates in preceding meal (g)	—	—	—	–0.0029
9. OGTT[e] fasting BG (mmol/L)	—	—	0.2974	—
10. OGTT 2h BG (mmol/L)	0.0484	0.0397	0.0356	0.1036
11. Fasting serum insulin (pmol/L)[f]	—	—	—	0,0021
12. Sports (≥2 days/week, yes/no)[g]	–0.1416	—	—	—
13. Climbing stairs (≥4 flights/day, yes/no)[g]	–0.0497	–0.1938	–0.1860	–0.0364
14. Walking (≤30, 31-60, ≥61 min/day for 0, 1, 2)[g]	—	–0.1062	–0.0864	–0.3349
15. Legumes >1/week ((yes/no)[g]	—	—	—	–0.2184
16. Coffee (0-1, 2-3, >3 cups/day for 0, 1, 2)[g]	0.0025	0.1173	0.0738	0.0311

[a]AUC60: area under the postprandial blood glucose curve 60 minutes after the meal.

[b]AUC120: area under the postprandial blood glucose curve 120 minutes after the meal.

[c]BG60: blood glucose level 60 minutes after the meal.

[d]BP: blood pressure. Peak BG: peak BG level on a 3-hour postprandial BG curve.

[e]OGTT: oral glucose tolerance test.

[f]Measured at the day of OGTT.

[g]During pregnancy.

Table 4. Estimation of model performance.

Characteristic and set	*R*	Root mean square error	Mean absolute error	Mean absolute percentage error
AUC60 [a]				
Test	.79	0.62	0.52	9.3%
Training	.78	0.51	0.40	6.8%
AUC120 [b]				
Test	.75	0.61	0.48	9.1%
Training	.75	0.51	0.39	6.6%
BG60 [c]				
Test	.69	0.81	0.66	12.0%
Training	.66	0.75	0.56	8.9%
Peak BG [d]				
Test	.48	1.00	0.77	12.2%
Training	.74	0.68	0.53	8.0%

[a]AUC60: area under the postprandial blood glucose curve 60 minutes after the meal.

[b]AUC120: area under the postprandial blood glucose curve 120 minutes after the meal.

[c]BG60: blood glucose level 60 minutes after the meal.

[d]BP: blood pressure. Peak BG: peak BG level on a 3-hour postprandial BG curve.

Discussion

Principal Results

Our infrastructure, including the mobile app for patients' electronic record management and data processing algorithms for matching of the CGM system signal and electronic diary, enabled the collection and analysis of data on 909 food intakes and corresponding postprandial BG curves from 62 pregnant women. Combining these data with patients' characteristics (eg, HbA_{1c}, BMI, age, and lifestyle parameters) facilitated the development of models that accurately predict PPG response for real-life meals and can be implemented in our app for personalized prediction of PPG response and subsequent decision-making support.

To the best of our knowledge, no previous study has developed a tool for personalized prediction of PPG response in pregnant women. As the physiology of pregnancy differs significantly from that of the nonpregnant state, models created on nonpregnant participants have no potential to be applied to the management of pregnant women.

A unique feature of our app is the ability to integrate food choices based on our PPG response models in a decision-making algorithm. This will enable the app to give personalized advice concerning each upcoming meal in order to achieve desired BG levels. To adapt the personalized model to her specific requirements, a woman simply needs to fill in a short questionnaire when the app is first started (the questionnaire covers parameters 9-16 listed in Table 3). In order to receive on-site personalized nutrition advice, participants have to enter the desired food components with exact weights before meals. The app will calculate the predicted PPG response and, in case the recommended levels are exceeded, will suggest how to reduce the carbohydrate content of the desired meal by reducing the amount of carbohydrate-rich products or by suggesting variants of products for replacement.

The significant potential of our app lies in the ability to capture data, provide decision support, and share data with health care providers, thus promoting communication. Although the efficacy in terms of use of our app is yet to be tested in an RCT, we believe that it already has potential for supplementing traditional care, especially between visits to the clinic, when patients can be provided with on-site personalized recommendations and education. Our app may also be helpful for the collection of large datasets for future statistical analyses.

For short-term use (eg, pregnancy), mHealth solutions might be more widely accepted and there will potentially be less app attrition compared to patients with chronic diseases.

To test and train PPG response prediction models, we used data of women with normal glucose tolerance in line with the data from GDM patients. We consider this approach appropriate because the postprandial BG levels of women with normal glucose tolerance and GDM diagnosed under the new IADPSG criteria often overlap, as shown in Table 2. Because women with normal glucose tolerance are not required to adhere to a diet, they consume much more carbohydrates. Consequently, they often have BG levels similar to GDM patients, thus enabling us to assess the impact on PPG response of products with a more variable carbohydrate content.

Our approach of making patients select foodstuff from a prebuilt database is a good way to standardize and ease the procedure of food tracking as well as to increase data reliability. On the other hand, this created an additional problem for patients who had to search for their desired food items.

We provided patients with the possibility to create their own food items by combining existing items, but only a few people found this feature useful. There are two main reasons why we refrained from allowing users to add food items to the app database on their own. Firstly, there is virtually no way to find a comprehensive list of all the nutritional characteristics we would like to track (previous 25 features) for potential food items. Secondly, there is a concern with the quality of data available on publicly available Internet sources. Allowing users to add new items into the database might lead to mistakes and incomplete data, which might in turn cause bias in the statistical analyses.

The most important predictors of PPG response that remained significant for all models were BG level before meal, quantity of carbohydrates in the meal, type of food (breakfast was a heightening factor), 2-hour BG level at OGTT, frequency of coffee consumption as a heightening factor, and the level of physical activity expressed as climbing stairs during pregnancy, which lowered PPG response. All these parameters are expected with regard to physiology and are in line with the results of previous studies addressing factors determining glucose metabolism [33-35].

Comparison With Prior Work

Table 5 compares the prediction quality of the developed predicting models to that of BG predicting models of different types presented in the literature. All models exhibit adequate accuracy that allows them to be used for patient assistance. The predictive power of the developed models and those presented in recent scientific papers is of equal quality in terms of *R*, RMSE, MAE, and MAPE. Considering that the developed model is simple, interpretable, and requires little time to tune its parameters, in which it fulfills the requirement for dynamical coefficient updates, the model must be stated as adequate for the pursued goal.

Limitations

Our study has several limitations. Because the data on food intake were self-reported by participants through the app, there is the potential for biases due to inaccurate reports. For example, women who are overweight or gain excessive weight during pregnancy often underestimate their actual consumption of foods that are considered harmful. This is a typical drawback of any epidemiological study assessing nutrition. However, we used hints found in the data to reduce this kind of bias, as described previously (see Control of the Accuracy of Self-Reports section).

In addition, the PPG response prediction models are based on data from a relatively small sample size. However, the resulting PPG response prediction models have predictive power of equal quality to those presented in recent scientific papers [10,36-39]. Overall, we believe that the developed PPG response prediction models are sufficiently accurate to form the basis for a subsequent self-management intervention.

There are several concerns limiting the current use of the developed app with the built-in decision-making algorithm. Our app is so far only available for Android devices and PCs; unfortunately, this prevents iPhone users from accessing the app using their mobile phone (they can access it only using their desktop computer). We plan to increase the number of potential users by implementing an app for iPhone users as well. We are also considering implementing a responsive Web app to allow usage of a wide range of platforms and form factors.

This study did not include insulin treatment. Thus, the current models are not designed to predict PPG response after insulin injections. This will need to be evaluated in the future.

Table 5. Comparison between the proposed model and prior developed models.

Value and author(s)	Mathematical model	*R*	Root mean square error (mmol/L*hour; mmol/L)	Mean absolute percentage error (%)
AUC120 [a]				
Pustozerov et al[b]	Linear regression	.71	0.62	6.8
Zeevi et al [10]	Boosted decision trees	.70	—	—
BG60 [c]				
Pustozerov et al[b]	Linear regression	.69	0.82	12.0
Plis et al [36]	Support vector regression	—	1.97	—
Wang et al [37]	Autoregression, support vector machines, and neural network	—	0.53-1.29	5.1-16.6
Perez-Gandia et al [38]	Neural network	—	1.38-1.60	—
Perez-Gandia et al [38]	Autoregression	—	1.67-2.17	—
Stahl [39]	Lehmann and Deutsch, Dalla Man	—	1.24-1.73	—

[a]AUC60: area under the postprandial blood glucose curve 60 minutes after the meal.

[b]Our model.

[c]BG60: blood glucose level 60 minutes after the meal.

Our app currently lacks the ability to wirelessly upload BG readings. We also plan to overcome this inconvenience in the future.

The algorithm used for data processing is currently only semiautomatic and requires the intervention of the researcher in the final data processing steps. In the future, all the system software should conceivably be rewritten in a high-level programming language.

The developed PPG response prediction models are based and validated on the data of women in the third trimester of pregnancy. Therefore, it should not be used in the first and second trimesters. However, GDM is usually diagnosed in the third trimester and, according to the recommendations of some diabetes associations, can be diagnosed only in the third trimester [8]. Because Russian guidelines suggest diagnosing GDM at any time during pregnancy [24], we are planning an additional study to validate the PPG response prediction models for the first and second trimesters.

To improve BG control, as well as maternal and fetal outcomes, the efficacy of this app for the management of women with GDM needs confirmation in an RCT and we plan to perform such an RCT.

Conclusions

This infrastructure comprising the data processing algorithms, the BG prediction models, and the mobile app for patients' electronic record management can be useful for guiding BG prediction-based personalized recommendations for GDM patients. The accuracy of the prediction models was validated on training sets of patients with 20-fold cross-validation for parameters tuning. The efficacy of the implementation in terms of providing health care to women with GDM to reduce BG levels and pregnancy complications will be evaluated in a future RCT.

Acknowledgments

This study was funded by the Russian Science Foundation (project no: 15-14-30012). We would like to thank Editage for English language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Lifestyle questionnaire.

[PDF File (Adobe PDF File), 23KB - mhealth_v6i1e6_app1.pdf]

Multimedia Appendix 2

Mobile app forms.

[MP4 File (MP4 Video), 2MB - mhealth_v6i1e6_app2.mp4]

Multimedia Appendix 3

CGMS and electronic diary data merging.

[MP4 File (MP4 Video), 4MB - mhealth_v6i1e6_app3.mp4]

References

1. HAPO Study Cooperative Research Group, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med 2008 May 08;358(19):1991-2002. [doi: 10.1056/NEJMoa0707943] [Medline: 18463375]
2. Silverman BL, Rizzo TA, Cho NH, Metzger BE. Long-term effects of the intrauterine environment. The Northwestern University Diabetes in Pregnancy Center. Diabetes Care 1998 Aug;21 Suppl 2:B142-B149. [Medline: 9704242]
3. El Hajj N, Schneider E, Lehnen H, Haaf T. Epigenetics and life-long consequences of an adverse nutritional and diabetic intrauterine environment. Reproduction 2014 Dec;148(6):R111-R120 [FREE Full text] [doi: 10.1530/REP-14-0334] [Medline: 25187623]
4. Bonoto BC, de Araújo VE, Godói IP, de Lemos LL, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
5. Liang X, Wang Q, Yang X, Cao J, Chen J, Mo X, et al. Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis. Diabet Med 2011 Apr;28(4):455-463. [doi: 10.1111/j.1464-5491.2010.03180.x] [Medline: 21392066]

6. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
7. Borgen I, Garnweidner-Holme LM, Jacobsen AF, Bjerkan K, Fayyad S, Joranger P, et al. Smartphone application for women with gestational diabetes mellitus: a study protocol for a multicentre randomised controlled trial. BMJ Open 2017 Mar 27;7(3):e013117 [FREE Full text] [doi: 10.1136/bmjopen-2016-013117] [Medline: 28348183]
8. American Diabetes Association. Standards of medical care in diabetes-2016 abridged for primary care providers. Clin Diabetes 2016 Jan;34(1):3-21 [FREE Full text] [doi: 10.2337/diaclin.34.1.3] [Medline: 26807004]
9. American Diabetes Association. (4) Foundations of care: education, nutrition, physical activity, smoking cessation, psychosocial care, and immunization. Diabetes Care 2015 Jan;38 Suppl:S20-S30. [doi: 10.2337/dc15-S007] [Medline: 25537702]
10. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, et al. Personalized nutrition by prediction of glycemic responses. Cell 2015 Nov 19;163(5):1079-1094 [FREE Full text] [doi: 10.1016/j.cell.2015.11.001] [Medline: 26590418]
11. Jenkins DJ, Wolever TM, Taylor RH, Barker H, Fielden H, Baldwin JM, et al. Glycemic index of foods: a physiological basis for carbohydrate exchange. Am J Clin Nutr 1981 Mar;34(3):362-366. [Medline: 6259925]
12. Dodd H, Williams S, Brown R, Venn B. Calculating meal glycemic index by using measured and published food values compared with directly measured meal glycemic index. Am J Clin Nutr 2011 Oct;94(4):992-996 [FREE Full text] [doi: 10.3945/ajcn.111.012138] [Medline: 21831990]
13. Vrolix R, Mensink RP. Variability of the glycemic response to single food products in healthy subjects. Contemp Clin Trials 2010 Jan;31(1):5-11. [doi: 10.1016/j.cct.2009.08.001] [Medline: 19737630]
14. Dunstan DW, Kingwell BA, Larsen R, Healy GN, Cerin E, Hamilton MT, et al. Breaking up prolonged sitting reduces postprandial glucose and insulin responses. Diabetes Care 2012 May;35(5):976-983 [FREE Full text] [doi: 10.2337/dc11-1931] [Medline: 22374636]
15. Carpenter D, Dhar S, Mitchell LM, Fu B, Tyson J, Shwan NA, et al. Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. Hum Mol Genet 2015 Jun 15;24(12):3472-3480 [FREE Full text] [doi: 10.1093/hmg/ddv098] [Medline: 25788522]
16. Gibbs EM, Stock JL, McCoid SC, Stukenbrok HA, Pessin JE, Stevenson RW, et al. Glycemic improvement in diabetic db/db mice by overexpression of the human insulin-regulatable glucose transporter (GLUT4). J Clin Invest 1995 Apr;95(4):1512-1518 [FREE Full text] [doi: 10.1172/JCI117823] [Medline: 7706456]
17. Oviedo S, Vehí J, Calm R, Armengol J. A review of personalized blood glucose prediction strategies for T1DM patients. Int J Numer Method Biomed Eng 2017 Jun;33(6):e1. [doi: 10.1002/cnm.2833] [Medline: 27644067]
18. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011;13(3):e65 [FREE Full text] [doi: 10.2196/jmir.1874] [Medline: 21979293]
19. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep;6(5):1197-1206 [FREE Full text] [Medline: 23063047]
20. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069 [FREE Full text] [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
21. Caburnay CA, Graff K, Harris JK, McQueen A, Smith M, Fairchild M, et al. Evaluating diabetes mobile applications for health literate designs and functionality, 2014. Prev Chronic Dis 2015;12:E61 [FREE Full text] [doi: 10.5888/pcd12.140433] [Medline: 25950568]
22. United States Department of Agriculture, Agricultural Research Service. USDA Food Composition Databases URL: https://ndb.nal.usda.gov/ndb/ [accessed 2017-12-08] [WebCite Cache ID 6vXpQKKDk]
23. Russian Academy of Medical Sciences. [Institute of Nutrition]. URL: http://www.ion.ru/ [accessed 2017-12-08] [WebCite Cache ID 6vXpfNtby]
24. Dedov II, Krasnopol'skiy VI, Sukhikh GT. Russian National Consensus Statement on gestational diabetes: diagnostics, treatment and postnatal care. DM 2012 Dec 15;15(4):4. [doi: 10.14341/2072-0351-5531]
25. International Association of Diabetes and Pregnancy Study Groups Consensus Panel, Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, et al. International Association of Diabetes and Pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care 2010 Mar;33(3):676-682 [FREE Full text] [doi: 10.2337/dc09-1848] [Medline: 20190296]
26. Durán A, Martín P, Runkle I, Pérez N, Abad R, Fernández M, et al. Benefits of self-monitoring blood glucose in the management of new-onset Type 2 diabetes mellitus: the St Carlos Study, a prospective randomized clinic-based interventional study with parallel groups. J Diabetes 2010 Sep;2(3):203-211. [doi: 10.1111/j.1753-0407.2010.00081.x] [Medline: 20923485]
27. IBM. SPSS Statistics, Version 22 URL: https://www.ibm.com/products/spss-statistics [accessed 2017-11-10] [WebCite Cache ID 6usdCz5np]
28. MathWorks. MATLAB and Statistics Toolbox Release 2016a URL: https://www.mathworks.com/products/statistics.html [accessed 2017-11-10] [WebCite Cache ID 6usdRluzy]

29. R Core Team. 2017 Nov 10. R: A language and environment for statistical computing, Release 3A URL: https://www.R-project.org/ [WebCite Cache ID 6usdb9lyb]
30. Python. Python Language Reference, Version 2 URL: https://www.python.org/ [accessed 2017-11-10] [WebCite Cache ID 6usdhLDTa]
31. Tibshirani R. Regression shrinkage selection via the lasso. J R Stat Soc B 1996;58(1):267-288. [doi: 10.1111/j.1467-9868.2011.00771]
32. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
33. Whitehead N, White H. Systematic review of randomised controlled trials of the effects of caffeine or caffeinated drinks on blood glucose concentrations and insulin sensitivity in people with diabetes mellitus. J Hum Nutr Diet 2013 Apr;26(2):111-125. [doi: 10.1111/jhn.12033] [Medline: 23331476]
34. Yamada Y, Hosoya S, Nishimura S, Tanaka T, Kajimoto Y, Nishimura A, et al. Effect of bread containing resistant starch on postprandial blood glucose levels in humans. Biosci Biotechnol Biochem 2005 Mar;69(3):559-566 [FREE Full text] [Medline: 15784985]
35. Schmidt MI, Hadji-Georgopoulos A, Rendell M, Margolis S, Kowarski A. The dawn phenomenon, an early morning glucose rise: implications for diabetic intraday blood glucose variation. Diabetes Care 1981;4(6):579-585. [Medline: 6751733]
36. Plis K, Bunescu RC, Marling CR, Shubrook J. A machine learning approach to predicting blood glucose levels for diabetes management. In: Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014 Jun 18 Presented at: AAAI Workshop: Modern Artificial Intelligence for Health Analytics; Jul 27-31, 2014; Québec City, QC p. 35-39 URL: https://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/view/8737
37. Wang Y, Wu X, Mo X. A novel adaptive-weighted-average framework for blood glucose prediction. Diabetes Technol Ther 2013 Oct;15(10):792-801 [FREE Full text] [doi: 10.1089/dia.2013.0104] [Medline: 23883406]
38. Pérez-Gandía C, Facchinetti A, Sparacino G, Cobelli C, Gómez EJ, Rigla M, et al. Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring. Diabetes Technol Ther 2010 Jan;12(1):81-88. [doi: 10.1089/dia.2009.0076] [Medline: 20082589]
39. Stahl F. Diabetes Mellitus Glucose Prediction by Linear and Bayesian Ensemble Modeling [PhD Thesis]. Lund, Sweden: Department of Automatic Control, Lund Institute of Technology, Lund University; 2012:127.

Abbreviations

AUC: area under the curve
AUC60: area under the postprandial blood glucose curve 60 minutes after the meal
AUC120: area under the postprandial blood glucose curve 120 minutes after the meal
BG: blood glucose
BG60: postprandial blood glucose 60 minutes after the meal
BMI: body mass index
CGM: continuous glucose monitoring
GDM: gestational diabetes mellitus
HbA_{1c}: glycated hemoglobin A_{1c}
MAE: mean absolute error
MAPE: mean absolute percentage error
mHealth: mobile health
OGTT: oral glucose tolerance test
Peak BG: peak blood glucose value on a 3-hour postprandial blood glucose
PG: plasma glucose
PPG: postprandial glucose
RCT: randomized controlled trial
RMSE: root mean square error

Edited by G Eysenbach; submitted 10.11.17; peer-reviewed by S Engeli, S Li, J Krebs; comments to author 06.12.17; revised version received 08.12.17; accepted 08.12.17; published 09.01.18.

Please cite as:
Pustozerov E, Popova P, Tkachuk A, Bolotko Y, Yuldashev Z, Grineva E
Development and Evaluation of a Mobile Personalized Blood Glucose Prediction System for Patients With Gestational Diabetes Mellitus
JMIR Mhealth Uhealth 2018;6(1):e6
URL: http://mhealth.jmir.org/2018/1/e6/
doi:10.2196/mhealth.9236
PMID:29317385

Original Paper

Young People's, Parents', and Professionals' Views on Required Components of Mobile Apps to Support Self-Management of Juvenile Arthritis: Qualitative Study

Jennifer M Waite-Jones[1*], BA (Hons), MSc, PhD; Rabiya Majeed-Ariss[2*], BSc (Hons), MSc, PhD; Joanna Smith[1*], BSc (Hons), MSc, PhD; Simon R Stones[1*], BSc (Hons); Vanessa Van Rooyen[3*], BSc (Hons); Veronica Swallow[1,4*], BSc (Hons), MMedSCi, PhD

[1]School of Healthcare, University of Leeds, Leeds, United Kingdom

[2]School of Psychological Sciences, University of Manchester, Manchester, United Kingdom

[3]Leeds Children's Hospital, Leeds, United Kingdom

[4]Faculty of Science, Charles Sturt University, Bathurst, Australia

[*]all authors contributed equally

Corresponding Author:
Veronica Swallow, BSc (Hons), MMedSCi, PhD
Faculty of Science
Charles Sturt University
Building 1293 Room 2
Bathurst,
Australia
Phone: 61 0419 490 431
Email: v.m.swallow@leeds.ac.uk

Abstract

Background: There is growing evidence that supporting self-management of Juvenile Arthritis can benefit both patients and professionals. Young people with Juvenile Arthritis and their healthy peers increasingly use mobile technologies to access information and support in day-to-day life. Therefore, a user-led, rigorously developed and evaluated mobile app could be valuable for facilitating young people's self-management of Juvenile Arthritis.

Objective: The objective of this study was to seek the views of young people with Juvenile Arthritis, their parents or carers, and health care professionals (HCPs) as to what should be included in a mobile app to facilitate young people's self-management of chronic Juvenile Arthritis.

Methods: A qualitative approach was adopted with a purposeful sample of 9 young people aged 10-18 years with Juvenile Arthritis, 8 parents or carers, and 8 HCPs involved in their care. Data were gathered through semi-structured focus group and individual interviews with young people and their parents or carers and HCPs. Interview discussion was facilitated through demonstration of four existing health apps to explore participants' views on strengths and limitations of these, barriers and facilitators to mobile app use, preferred designs, functionality, levels of interaction, and data sharing arrangements. Data were analyzed using the framework approach.

Results: Analysis revealed three interlinked, overarching themes: (1) purpose, (2) components and content, and (3) social support. Despite some differences in emphasis on essential content, general agreement was found between young people with Juvenile Arthritis their parents or carers, and professionals that a mobile app to aid self-management would be useful. Underpinning the themes was a prerequisite that young people are enabled to feel a sense of ownership and control of the app, and that it be an interactive, engaging resource that offers developmentally appropriate information and reminders, as well as enabling them to monitor their symptoms and access social support.

Conclusions: Findings justify and pave the way for a future feasibility study into the production and preliminary testing of such an app. This would consider issues such as compatibility with existing technologies, costs, age, and cross-gender appeal as well as resource implications.

(JMIR Mhealth Uhealth 2018;6(1):e25) doi:10.2196/mhealth.9179

KEYWORDS

Adolescent; young people; juvenile arthritis; mobile apps; self-management; qualitative research

Introduction

The self-management support needs of young people with long-term conditions (LTCs) such as juvenile arthritis are now better understood [1]. They include the following core skills: problem solving, decision making, resource utilization, and patient and professional relationship formation [2]. Learning the skills and knowledge needed to manage treatment regimens effectively is a key component of LTC management for young people with Juvenile Arthritis as they grow up and transition to adult health care [3-6]. Acquisition of such skills is particularly important during this period of transition, given the changes within previous support structures which included parents or carers and pediatric focused health care professionals (HCPs). Foster et al [7] propose that professionals should encourage young people to take a lead role in managing their condition while supporting parents or carers in the changes to their previous roles and responsibilities.

Stinson et al [8] describe self-management as taking responsibility for symptoms and treatment of a chronic condition, including the ability to cope with the psychological and physical changes and the effects of adopting a more appropriate lifestyle. Coping methods involved gaining information, learning appropriate behavior changes, and improving self-efficacy.

Improved self-efficacy is essential to the ability to be able to self-manage a chronic condition. The concept of self-efficacy can be understood in terms of social cognitive theory developed by Bandura [9,10], which suggests that to successfully promote positive behavior change, there is a need to encourage individuals to gain mastery over their situation through verbal persuasion, improved communication, and vicarious learning from role models. To improve self-efficacy, according to Schwartz et al [3], reciprocal relationships need to be encouraged between patients, parents or carers, and professionals to foster good communication and the motivation for patients to achieve appropriate goals and improve knowledge. Elwyn et al [11] suggest that through sharing decision making within reciprocal relationships, supporting patients, and helping them make decisions based on informed preferences promotes successful self-management.

Furthermore, facilitation of such learning must be carefully tailored to young people's needs, reflecting a growing evidence base, which suggests that brain development continues up until a young person is in their early twenties [12]. Therefore, young people need support and guidance to become accountable for their own health, enabling them to develop into independent, empowered, and responsible adults [7].

The Internet is becoming the preferred method for young people when seeking health information [13], particularly when searching for tools enabling self-management of chronic conditions [13]. Web-based self-management strategies have been reported by young people to be *a promising avenue* to support them in managing their symptoms and control their emotions as well as access information and social support [8]. Mobile apps can also provide information and integrate health care tasks [14]. Therefore, self-management tools provided by mobile apps could help meet the requirements of young people with Juvenile Arthritis and those involved in their care [15].

Such apps need to be engaging and have usability as well as functionality [16]. The use of mobile apps to improve pain management in young people with Juvenile Arthritis has been studied [17], but there is a lack of rigorously developed apps to support other aspects of self-management, and this suggests a need for further evaluation through rigorous research to ensure best practice [18].

This study completes stage three of the modeling stage [19] of development and evaluation of a mobile app for self-management by young people with Juvenile Arthritis. The first stage involved a user-led Web-based survey of 14 young people with juvenile arthritis (12 females and 2 males aged between 11 and 22 years) to assess the need for a self-management mobile app. A total of 12 respondents felt there was a need for the mobile app and 2 were unsure. Components specified included the following: appointment and treatment reminders, a tool to record/monitor disease activity and quality of life, a goal-setting facility, and a secure central point to share experiences with peers. Places where respondents said they would use such an app included the following: home, school, shopping, hospital, and when with friends [20].

Stage two included a systematic review of the effectiveness of apps for self-management by young people with physical LTCs. Few studies were found to use a systematic, young-person and family-led approach to app development and evaluation, thus confirming the need for this study [14].

The third stage included a qualitative study to understand the preferred components of a mobile app for young-people with juvenile arthritis, their parents or carers, and professionals before developing an app.

The objective of this study was to seek the views of young people with Juvenile Arthritis, their parents or carers, and HCPs as to what should be included in a mobile app to facilitate young people's self-management of chronic rheumatic disease.

Methods

Design

This study was informed by previous stages of the project as well as an additional electronic scoping exercise guided by the Cochrane Handbook [21]. A qualitative design was adopted, which investigated a potential aid to self-management of young people's Juvenile Arthritis based on the concept of self-efficacy [9,10], and the Medical Research Council [19] framework for the development and evaluation of complex interventions was followed.

Ethics

Ethical approval was obtained from the National Health Service Health Research Authority (reference no: 193786).

Participants

Purposeful sampling of young people aged 10-18 years from the database of a pediatric rheumatology clinic based in a large teaching hospital in the north of England catering for different forms of Juvenile Arthritis was carried out by a rheumatology nurse specialist (VVR). To be eligible for recruitment, participants sought had to have been diagnosed with a chronic rheumatic disease. Although the composition of the group depended upon those who agreed to participate, an attempt was made to achieve variation regarding age, developmental stage, disease type and duration, ethnicity, sex, socioeconomic status, and treatment type.

VVR also invited parents or carers and professionals caring for these young people to participate. The final sample group comprised 25 participants, including 9 young people, 8 parents or carers, and 8 HCPs (Table 1). The HCP group comprised consultants, clinical nurse specialists, a psychologist, a youth worker, and a pharmacist.

Verbal and written information (in the form of approved developmentally appropriate information sheets, topic guides, and consent or assent forms) was offered to potential participants. Participants aged 16 years and older provided VS and JW-J with written consent. Patients under 16 years signed assent forms, and their parents or carers signed consent forms on their behalf [22].

Table 1. Participants' demographics. HCP: health care professional; PC: parents or carers; YP: young people.

Participant and participant number	Sex	Age (years)
Young people (N=9)		
YP1	Female	17
YP2	Male	14
YP3	Female	13
YP4	Female	11
YP5	Female	14
YP6	Female	13
YP7	Female	10
YP8	Male	15
YP9	Female	15
Parents or carers (N=8)		
PC1	Female	N/A[a]
PC2	Female	N/A
PC3	Male	N/A
PC4	Female	N/A
PC5	Male	N/A
PC6	Female	N/A
PC7	Female	N/A
PC8	Female	N/A
Health care professionals (N=8)		
HCP1	N/A	N/A
HCP2	N/A	N/A
HCP3	N/A	N/A
HCP4	N/A	N/A
HCP5	N/A	N/A
HCP6	N/A	N/A
HCP7	N/A	N/A
HCP8	N/A	N/A

[a]N/A: not applicable.

Tools

Developmentally appropriate topic guides (see Multimedia Appendix 1) were used to structure focus groups and individual interviews [23]. Four existing self-management apps were also demonstrated. Two of these were designed for adults with rheumatoid arthritis [24,25], the third was for adults with chronic pain [26], and the fourth was aimed at younger people with type 2 diabetes mellitus [27]. These particular apps were selected on the advice of the user ambassador and coapplicant (SS) (in line with our philosophy of an evidence-based user-led approach, as recommended within our recent systematic review [14]) who thought they offered specific systems relevant to Juvenile Arthritis.

Data Collection

Using a participatory approach [28], for participants' convenience, semi-structured interviews were conducted with young people and their parents or carers, and information gathered from the professionals via 2 focus groups. Although offered the opportunity to be interviewed separately from their parent or carer, only one young person expressed a desire for an independent interview and this was only because they were available at a different time to their parents. Developmentally appropriate topic guides were used to explore participants' information needs, experience of using mobile apps, and opinions of relevant current mobile apps. The sample apps were demonstrated to generate discussion around participants' views on strengths and limitations of existing mobile apps, barriers and facilitators to mobile app use, preferred designs, functionality, levels of interaction, and data sharing arrangements. Participants were then asked for additional comments. On completion of interviews, the young participants were sent a £10 thank you voucher, and the downloaded information from support groups was posted to young people and parents or carers. Focus groups and interviews lasted between 35 and 60 min and were digitally recorded and transcribed by the first author.

Data Analysis

Data were analyzed by 5 members of the research team and included 2 children's nurse researchers and 2 psychologists specializing in children's health who all have considerable experience in different forms of analysis used in qualitative studies. In addition, the user ambassador and fellow researcher was also involved in each stage of the research project, including analysis. The framework approach was used as an analytic method as this flexible, systematic, and rigorous method offers clarity, transparency, and an audit trail [29-31]. Two transcripts were initially thematically coded by 5 members of the research team, and 4 overarching themes with related subthemes emerged. These provided a framework through which all the transcripts were analyzed and displayed via a bespoke matrix in Microsoft Excel [32]. Once all transcripts had been analyzed, themes were discussed, refined, and critically evaluated resulting in 2 of the original themes being integrated. All final themes, subthemes, and relevant quotations were then reviewed by the whole team and a consensus was achieved that they clearly reflected participants' views.

Results

Findings were grouped into 3 interlinked, overarching themes: (1) purpose, (2) components and content, and (3) social support (Textbox 1).

Textbox 1. Core themes and subthemes.

Purpose

- App ownership
- Monitoring chronic rheumatic disease (Juvenile Arthritis) and information sharing
- Facility for reminders

Components and content

- Desired components
- Essential content
- Practical considerations

Social support (emotional and practical)

- Access and signposting to existing support networks
- Secure peer support
- Understanding from others without Juvenile Arthritis
- Parent support network

Theme 1: Purpose of Mobile Apps

App Ownership

Young people stressed the importance of app ownership and being able to choose if and how they would use it:

> *Something that you don't have to get from your doctor...You can use it or not use it...not be forced to.* [YP1, young person #1]

Overall, a personalized approach was preferred so that individuals could feel supported and empowered by the app:

> *Yes, it is like there is action. It is like "Here we are here for you!"* [YP5]

Coded access was suggested by young people and parents or carers to ensure the app provided a secure, supportive environment:

> *So that people who do not have the code and do not have arthritis can't get on, so it is as easy as that.* [YP1]

Although professionals agreed that some form of secure access was needed, they also commented on *good* uses of the app and expressed some concern relating to young people's desires for ownership:

> *We would have to be careful because it only takes someone with an app joining a support group and they can easily access cards, you know things that they can use and have an understanding of the system to manufacture a diagnosis of arthritis...We have come across a lot of kids who have said that they have arthritis that it is clear that they don't.* [HCP2]

Monitoring Juvenile arthritis and Information Sharing

Most participants thought that monitoring Juvenile Arthritis and pain levels would be useful for young people, as illustrated by the following response to the demonstration of an existing app:

> *That tracking your pain thing [is useful] so that you can monitor [pain].* [YP7]

And the suggestion that:

> *You could have a little diary in there as well so if you want to write about your day, because if it helps you and you prefer to write, then maybe you could just type it in and it monitors it.* [YP4]

However, a few young people and parents or carers preferred not to focus on pain if it is low-level:

> *I do not think that I would like that app as much because it is too pain orientated...implying that you are in constant pain...if you were just in minor pain and you were more interested in other things.* [YP8]

> *And I think that this might be a little bit too much adult...because she knows when she is in pain...I don't know whether she would do that [use pain tracker].* [PC1, parent or carer #1]

In contrast, professionals mostly discussed young people's condition monitoring in the context of information sharing to enable efficient use of clinic appointments. For example, an app was viewed as a better alternative to the current paper-based option for young people to report on the frequency or level of pain, and medication use, between appointments:

> *So, apps can certainly be used to gain and to assess information and bring it to clinic...also feedback in real time.* [HCP2]

Young people discussed information-sharing in the context of a mutual exchange of communication with professionals, suggesting for example:

> *Your doctor can write something in [the app] maybe that you can read. Or...automatically send some leaflets or something to your house.* [YP4]

Young people explored the usefulness of parents' or carers' access to the app. One young person suggested that, depending on their age, young people could have the option when entering data to decide whether to share it with their parents or carers:

> *So, the younger ones [children], their parents can look.* [YP1]

Facility for Reminders

All participants recommended a facility for sending reminders, for example, to take medication and attend hospital appointments. One young person suggested a medication reminder that "[c]ould remind you again a couple of days later" [YP5].

A parent or carer suggested:

> *You would have to get the appointment somehow, networked to the unit at the hospital for them to be able to put the information in.* [PC3]

Additionally, professionals recommended adding reminders to the app about issues they would normally put in a letter to the family:

> *You have a clinic appointment so please bring shorts, or dress appropriately...that I suppose will help us a lot.* [HCP3]

Theme 2: Components and Content of Mobile Apps

Desired Components of the App

Professionals referred to existing apps that could be adapted for use by young people with Juvenile Arthritis. They discussed the usefulness of components such as mindfulness and relaxation techniques, as well as tools, to monitor pain and report treatment side effects. In addition, tools aimed at improving treatment adherence, such as schedules for regular investigations, were considered important for optimal management of Juvenile Arthritis. However, these were generally not seen as important by young people:

> *It is something that I would look at and then just...move it to one side.* [YP1]

Professionals, young people, and parents or carers unanimously recommended gamified apps. Professionals thought apps needed to be fun and interesting to help young people learn about their Juvenile Arthritis. Young people also favored quiz-styled

learning activities and characters they could *adopt* as part of the app, such as a pet:

> *If you made it into a game like, if you had a character to help you learn it more and help you and stuff you could make it give you information and then you could feed and bath your pet and stuff. And, it answers your questions and it gives you like notifications every day to feed it and wash it and stuff and you can put your pain scales in and stuff like that. And, as well it takes your mind off of your pain as well.* [YP7]

Although one professional said that gamified apps can be suitable for all age groups, young people emphasized that it should be age-appropriate. Professionals also suggested that young people could receive digital currency or rewards for self-management activities, such as taking their medication:

> *Because there is no incentive to do apps unless you get a reward.* [HCP3]

Several young people favored competitive behavior through a gamified app, suggesting a leader board as an incentive to encourage others to use the app in self-management. However, professionals were concerned about introducing competition, as this could encourage young people to compare themselves with others, overlooking the unique nature of their conditions as:

> *You run the risk of them comparing themselves to someone else who is doing much better than they are.* [HCP8]

Essential Content

Information provision was viewed as an essential feature of an app specific for young people with Juvenile Arthritis. Information about the condition, treatments, symptom management, and how other young people are living with the condition were identified as relevant content. However, it was stated that information should be presented in an age-, gender-, and developmentally appropriate way, tailored for different reading abilities and learning styles and in the form of fun activities. For example, the professionals pointed out that:

> *It is not all about saving our time though and taking away things from what we are doing, it's about engaging young people.* [HCP4]

All participants recommended that the style and format of presentation of materials be visually appealing to promote engagement with the app. Examples of ways to present information included *bite-size chunks* that could be built on video clips. The visual impact and interactivity was highly important to young people:

> *Something that is fun, engaging, interactional but educational. Fun but you learn something too...us teenagers, like we are quite picky and it's the look...so if it's just writing then it will be oh just, you know flick it through and get bored.* [YP1]

Practical Considerations of Apps

Several practical considerations were raised by respondents in relation to using the app. Safeguarding was an important issue for all participants, and it was suggested that the app was password protected, with professionals providing young people with the password. Requirements such as updating the app regularly and compatibility with new and existing mobile technology were expressed. For example, one parent or carer explained that:

> *I have had games in the past which will suddenly stop working and I know that it is the system because they have been updated past the point that, that system can cope with.* [PC3]

One young person suggested that the app contents should be backed up to a cloud storage facility so that the information could be downloaded to multiple devices with the appropriate permissions. Being able to access the app any time and across devices, with or without Wi-Fi or cellular data, was seen as important by young people and parents or carers, as not all young people have access to Wi-Fi.

All participants recognized the costs associated with developing an app, and although the availability of a free app was preferable, parents or carers and young people suggested that if there was a charge, a free trial was essential to enable them to self-assess whether the app was beneficial or not:

> *I think that you need to try before you buy. If you think...this is fantastic, it does not matter how much it is at that point, you will put your money into it because it works for you.* [PC3]

One young person suggested that existing reviews and comments from users would influence their decision to use an app. However, parents or carers' views differed, with one individual suggesting that they would prefer to either pay upfront, or not at all. Professionals raised important issues about commissioning of the app within the current health care infrastructure. HCP2 referred to the need for discussion with commissioners regarding tariffs, particularly when considering the long-term use and applicability of such apps.

Theme 3: App-Enabled Social Support (Emotional and Practical)

Access or Signposting to Existing Networks

All participants felt that an app should provide some form of social support. Professionals suggested signposting app users toward existing voluntary agencies offering information and support, such as:

> *JIA NRAS [Juvenile Idiopathic Arthritis, National Rheumatoid Arthritis Society] but it is for parents really, Arthritis Research UK and Arthritis Care.* [HCP7]

Several parents or carers and young people also thought signposting to social events would be important for young people as:

> *Then you can tell your parents about it and so maybe it could go through to your parents as well.* [YP4]

Secure Peer Support

Almost all young people suggested that the app should provide contact with others diagnosed with Juvenile Arthritis. For example, one young person explained that:

> *I didn't really know anybody at all that had arthritis.* [YP1]

Young people consistently saw this as a way of sharing experiences as:

> *It would be really good to see what other people have gone through and if I can relate to them [YP6], ask: How do you guys cope? [YP1] and: Feel kind of welcome to this little group of ours.* [YP5]

Young people expressed a need for contact with others with the same condition, despite help given by parents or carers and professionals:

> *You know that they are really trying to help you but it makes you frustrated especially because they are not going through what you are.* [YP4]

Professionals and parents or carers also saw the value of peer support. For example, a professional pointed out that it "can be really helpful" [HPC8].

And one mother explained that her daughter:

> *Would know a lot more about what she has to face if she could talk to other children who have arthritis.* [PC4]

Professionals and parents or carers recommended a secure, moderated social network that young people could opt into with consent as "Many [with Juvenile Arthritis] feel isolated" [HCP3] and "Just being able to talk to someone is sometimes quite a big thing" [PC6]. All young people saw a chat room as a useful way to access peer support: "[a] group chat, that would be amazing" [YP1]. YP2 said, "If I am like getting worried about it (Juvenile Arthritis), I might go in and ask someone".

Even a young person who firmly stated that they would not contribute to a chat room added, "But I could see like everybody else" [YP8].

However, it was recognized that: "A chat room would need to accommodate the different age groups" [PC4].

Issues relating to safeguarding emerged with young people raising security concerns, such as:

> *If somebody who did not actually have it [arthritis] but just pretended to have it to talk to children, well that could be a problem.* [YP7]

Repeated reference was made to the need for secure, moderated access as a way of restricting inappropriate users, although "Extra resources would be needed to maintain this" [PC5].

One parent or carer suggested the safer alternative of a message board or forum that enabled young people and parents to post and receive responses in real time and "could be easily monitored...by an administrator" [PC3].

Understanding From Others Without Juvenile Arthritis

Some means of educating peers about Juvenile Arthritis was requested by young people and parents or carers. For example, one young person explained that when taking tablets in the company of peers, everybody looked away so:

> *Something as depressing and serious as taking medication should be put in [the app]…it would be very useful for a lot of people including me.* [YP6]

Even professionals were seen to require more understanding as:

> *He [physio] said it is all up here, it is up in your brain...pain is an everyday thing.* [YP1]

Information provision in the form of an app was also viewed as a potential route for young people and their parents or carers to effectively explain the condition to other agencies, for instance the young person's school:

> *You then could show them [teachers and peers] this app and that this kind of medication, sometimes it can do A, B and C to me.* [YP5]

Parent Support Network

That the app could also include some form of access and support for parents was referred to by parents or carers as they too could feel isolated, having only spoken to professionals:

> *I have never spoken to a mum or a dad [about Juvenile Arthritis, ever.* [PC6]

They explained that:

> *Some parents can find it more stressful to deal with than others.* [PC4]

So through such access:

> *You can share information and ideas with other parents as to how you have dealt with it.* [PC4]

Young people also recognized this need and suggested that:

> *You could have a little chat room between parents because it is hard work.* [YP9]

And it would be:

> *Something for parents in addition to young people so they can monitor and support their child.* [YP1]

Discussion

Principal Findings

Findings from this study suggest that the purpose of an app for self-management of Juvenile Arthritis should be to provide young people with the ownership and control of an interactive, engaging tool that gives information, monitors symptoms, offers reminders, and provides social support. Using the different elements of the app to manage their condition potentially enables young people with Juvenile Arthritis to develop a greater sense of autonomy through sharing the responsibility of managing their condition with professionals, as advocated by Foster et al [7], Schwartz et al [3], and Stinson et al [8]. Sharing access with professionals and discussing experiences recorded on the app

can also help in developing good relationships between patients and professionals. Therefore, such an app will help young people with Juvenile Arthritis to develop the essential core skills of problem solving, decision making, resource utilization, and relationship formation with professionals, as identified by Lorig and Holman [2].

Cai et al [33] and Whitehead and Seaton [34] report similar findings to those emerging from this study and stress how such electronic devices have the potential to offer convenient ways of providing self-management interventions. Analysis of data from this third phase of our project also confirms the conclusions drawn from our previous survey [20] and supplements the findings of our previous systematic review [14]. Moreover, they meet the criteria currently established by Wyatt and Williams [35] that such an app should be evidence-based, up-to-date, interactive, age-specific, and include privacy precaution to promote self-management.

Young people's request for an age-appropriate app with the potential for users to control the presentation also reflects the need for customization found within previous studies, such as that by Cai et al [33], Kenny et al [36], and Liptzin and Szefler [37]. The need for interactivity with potential rewards for good self-management to promote engagement identified by participants has also been reported by Cai et al [33]. This confirms the usefulness of games for self-management of chronic conditions [38-40], and the need to heed the advice of et al [41], who stress the importance of further work to identify potential gaming trends of the future. Such future work is important as the style of the game must be carefully designed because, although gamified apps are consistently downloaded by young people, inclusion of such facilities does not appear to influence the way many apps are rated in terms of popularity [41].

Some practical challenges were identified by participants including cost, compatibility with other technology, and the need to constantly update some of the content. Such concerns reflect those expressed in previous research, particularly given how the speed of technological advances may mean that modifications could be outdated even during development and evaluation and changes in equipment and functionality create difficulties when designing a mobile app [16].

That some form of access to peers should be included in an app has been found in previous studies, such as those by Cai et al [33] who recommend this should be accessed through signposting to existing organizations such as voluntary groups. Although such signposting was identified as useful in this study, there was considerable desire expressed by young people for the app to provide links with others diagnosed with Juvenile Arthritis through some form of chat room. Indeed, the number of comments offered by young people and parents or carers requesting peer-based social support was so noteworthy within this study that a second paper is planned to explore why this need is still so great, despite the support currently available from rheumatology clinics and voluntary groups. Only one young person did not see peer support as an important function of an app, but even this participant saw the value of observing how others were coping with their illness. Nevertheless, young people's request for a chat room facility raised challenges related to safeguarding and monitoring contributions, also recognized by parents or carers and professionals, and requiring serious consideration when designing an app.

Two novel findings of this study are: the suggestion that an app could be used to help foster understanding from others without Juvenile Arthritis and accelerate the process of informing other agencies such as schools and social services and the request for a parent support network to be included in the app. Acknowledging input from parents could prove fruitful, particularly because they are still the most popular source of health information for young people [13].

Limitations

Adopting a qualitative approach within this study has meant that it has some limitations. The limited opportunity within sampling to fully represent the number of young people in terms of age, gender, disease type, severity, longevity, and geographical location must be acknowledged. This was a single site study and for pragmatic reasons, it was only possible to invite young people diagnosed with vasculitis, uveitis, and JIA (systemic, polyarticular, and oligoarthritis) to participate. Moreover, it was only possible to invite 2 young males (10 to 12 years) from an ethnic minority background. One of these young males declined to take part, and the other did not complete the recruitment process. In addition, it was only possible to interview 2 of the 6 eligible male participants. Reasons given for not finally taking part in interviews included loss of interest and flare-up of illness.

Larger follow-up studies would enable inclusion of more males, ethnic minority groups, and other long-term conditions. This would be useful in case any cultural or gender differences could be identified which would supplement the findings reported here. Interestingly, Weiser [42] reported that young males use electronic media more for leisure and entertainment compared with young females who rather use it for educational purposes and interpersonal communication. Moreover, the participants of a study by Kenny et al [36] suggested that apps may provide an emotional outlet for males who may normally be reticent to display emotions.

However, the strength of this study is that it is part of a phased approach to development and evaluation of a complex intervention in line with the Medical Research Council framework [19]; it gathers the views of users, adopts a team approach, and includes user representation at all stages. Gathering data from interviewing young people and parents or carers together provided rich, detailed insight into their views about mobile apps and was not felt to inhibit contributions from younger participants. Indeed, it was noted that in each case, parents or carers consciously *took a back seat* and only offered information when asked, or prompted a young person when appropriate. It was particularly noteworthy just how the young people and their parents or carers enthusiastically encouraged each other in imagining different forms of informative games that could be included in the app.

Including young people and their parents or carers and professionals provided different perceptions of how mobile apps

can aid self-management of Juvenile Arthritis. Although there was much agreement between the three groups, on some occasions, there was a difference in the degree of emphasis placed on some features of such an app. For example, professionals placed great emphasis on how shared information from an app could save time and staff engagement during consultations; however, young people and their parents or carers saw this facility more as allowing communication in the periods between consultations. Moreover, although peer support was a priority for young people, it was only mentioned briefly by professionals who may have already considered some of the concerns this may raise if included in the app.

Interestingly, although all professionals and parents or carers saw pain management as an essential feature, there were differences here between the views of young people. Some wanted pain monitoring, but others did not want this to be a dominant feature of the app as they did not want to be reminded of pain when not actually experiencing it. This has implications for the use of current pain apps by young people. It suggests that when including this feature in a self-management app, its usefulness be made clear, particularly regarding recording pain levels during the past and previous days and not just the current time. The opportunity to identify differences in the views between those treating and those living with Juvenile Arthritis that was enabled by this study, confirms the need for patient and HCP partnerships in all areas of care and health care research.

The completion of this modeling stage as a precursor to developing and evaluating an app paves the way for a future full trial that will include patient and professional representation within a team approach.

Implications for Practice

Initially, introducing such an app could place more demands on health professionals treating those with Juvenile Arthritis. However, in the longer term, it is likely to alleviate their workload and improve communication and patients' well-being and potentially influence a smoother transition process as young people will be better engaged in their health care. Rigorous, comprehensive commissioning for professional involvement and app moderation needs to be considered at the outset of app development.

Conclusions

Agreement was found between young people with Juvenile Arthritis, their parents and carers, and professionals that a mobile app aiding self-management would be useful. This was provided that young people held a sense of ownership and control of an interactive, engaging device offering information and reminders as well as monitoring symptoms and providing social support. Findings justify and pave the way for a future feasibility study into the production and preliminary testing of such an app.

Acknowledgments

The authors would like to thank the participants who provided data for this study. They would also like to acknowledge the advice given by Dr Janet McDonagh (University of Manchester), Dr Mark Wood, (Leeds Children's Hospital), Mr Andrew Hall (University of Manchester), and Sally Vincent (Arthritis Care) when planning this study.

Funding for conducting this study was competitively awarded through the University of Leeds Pump Priming Programme.

Authors' Contributions

JMWJ and VS gathered data by means of recorded focus groups. JMWJ gathered data by means of individual interviews and transcribed all the interview recordings. JMWJ, VS, RMA, JS, and SRS analyzed the data. All authors contributed to drafts and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Topic Guides.

[PDF File (Adobe PDF File), 102KB - mhealth_v6i1e25_app1.pdf]

References

1. Health for the world's adolescents: A second chance in the second decade. In: World Health Organisation. Geneva: World Health Organisation; 2014:A.
2. Lorig K, Holman H. Ann Behav Med. Self-management education: History, definition, outcomes, and mechanisms. Ann Behav Med; 2003:1.
3. Schwartz L, Tuchman L, Hobbie WJ. A social-ecological model of readiness for transition to adult?oriented care for adolescents and young adults with chronic health conditions. Child: care, health and development; 2011:883-895.
4. Gilleland J, Amaral S, Mee LR. Journal of Pediatric Psychology. Getting ready to leave: Transition readiness in adolescent kidney transplant recipients. Journal of Pediatric Psychology; 2012:85-96.

5. Colver A, Longwell S. New understanding of adolescent brain development: relevance to transitional healthcare for young people with long term conditions. Arch Dis Child 2013 Nov;98(11):902-907 [FREE Full text] [doi: 10.1136/archdischild-2013-303945] [Medline: 23986559]
6. Stinson J. International Journal of Adolescent Medicine and Health; (6). 2013. A systematic review of transition readiness and transfer satisfaction measures for adolescents with chronic illness URL: https://www.degruyter.com/view/j/ijamh.2014.26.issue-2/ijamh-2013-0512/ijamh-2013-0512.xml [accessed 2018-01-16] [WebCite Cache ID 6wVynreKQ]
7. Foster HE. EULAR/PReS standards and recommendations for the transitional care of young people with juvenile-onset rheumatic diseases. Annals of the Rheumatic Diseases: p. annrheumdis-2016-21. Doi.org/10.1136/annrheumdis-2016-21 2016:0112. [doi: 10.1136/annrheumdis-2016-210112]
8. Stinson JN, Toomey PC, Stevens BJ, Kagan S, Duffy CM, Huber A, et al. Asking the experts: exploring the self-management needs of adolescents with arthritis. Arthritis Rheum 2008 Jan 15;59(1):65-72 [FREE Full text] [doi: 10.1002/art.23244] [Medline: 18163408]
9. Bandura A. Self-efficacy: The exercise of control. New York: W H Freeman; 1997.
10. Bandura A. Health promotion from the perspective of social cognitive theory. Psychology & Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
11. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
12. Sawyer SM, Afifi RA, Bearinger LH, Blakemore S, Dick B, Ezeh AC, et al. Adolescence: a foundation for future health. The Lancet 2012 Apr;379(9826):1630-1640. [doi: 10.1016/S0140-6736(12)60072-5]
13. Wartella E, Rideout V, Montegue H, Beaudoin-Ryan LA. Teens, health and technology: A national survey. Media and Communication 2016;4(3):A. [doi: 10.17645/mac.v4i3.515]
14. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and Adolescents: A Systematic Review of Adolescents' Use of Mobile Phone and Tablet Apps That Support Personal Management of Their Chronic or Long-Term Physical Conditions. J Med Internet Res 2015 Dec;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
15. Anderson K, Burford O, Emmerton L. App Chronic Disease Checklist: Protocol to Evaluate Mobile Apps for Chronic Disease Self-Management. JMIR Res Protoc 2016 Nov 04;5(4):e204 [FREE Full text] [doi: 10.2196/resprot.6194] [Medline: 27815233]
16. Hides L. 36(3). Are SMARTapps the future of youth mental health? InPsych: The Bulletin of the Australian Psychological Society Ltd, 2014 URL: http://search.informit.com.au/documentSummary;dn=685995600257622;res=IELAPA> [accessed 2018-01-16] [WebCite Cache ID 6wW1RIdtW]
17. Stinson J, White M, Isaac L, Campbell F, Brown S, Ruskin D, et al. Understanding the information and service needs of young adults with chronic pain: perspectives of young adults and their providers. Clin J Pain 2013 Jul;29(7):600-612. [doi: 10.1097/AJP.0b013e31826dce65] [Medline: 23328333]
18. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
19. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655. [doi: 10.1136/bmj.a1655]
20. Stones S, Majeed-Aris R, Hal A, Thomson W, Foster H, Baildam E, et al. Developing mobile applications for and with young people with long-term conditions. Poster. November 2014. [doi: 10.13140/RG.2.1.2761.4566]
21. Higgins JS. Vol. Cochrane handbook for systematic reviews of interventions URL: http://training.cochrane.org/handbook [accessed 2018-01-16] [WebCite Cache ID 6wW1TWMUg]
22. Gilljam B, Nygren J, Arvidsson S, Svedberg P. THU0610-HPR Involving Children with Juvenile Idiopathic Arthritis in Health-Related Research – Why and How? Ann Rheum Dis 2015 Jun 09;74(Suppl 2):1312.2-131312. [doi: 10.1136/annrheumdis-2015-eular.2044]
23. Huang X, O'Connor M, Ke L, Lee S. Ethical and methodological issues in qualitative health research involving children: A systematic review. Nurs Ethics 2016 May;23(3):339-356. [doi: 10.1177/0969733014564102] [Medline: 25552586]
24. Arthritis Foundation (2017). Track+React URL: http://www.arthritis.org/living-with-arthritis/tools-resources/track-and-react/ [accessed 2018-01-16] [WebCite Cache ID 6wW1VQPhc]
25. National Rheumatoid Arthritis Society (2017). Know your DAS: App (mobile and tablet) URL: https://www.nras.org.uk/know-your-das-mobile-app [accessed 2018-01-16] [WebCite Cache ID 6wW1XeaWs]
26. Paintoolkit. The Pain Toolkit. 2017. The Pain Toolkit URL: https://www.paintoolkit.org/ [accessed 2018-01-16] [WebCite Cache ID 6wW2IPNsI]
27. mySugr. 2017. URL: https://mysugr.com/ [accessed 2018-01-16] [WebCite Cache ID 6wW264HT6]
28. Groundwater-Smith S, Dockett SD. Participatory research with children and young people. London: Sage; 2014.
29. Smith JFJ. Nurse Res. 2011. Qualitative data analysis: the framework approach URL: http://eprints.hud.ac.uk/id/eprint/18884/1/RepositoryFrameworkNResearcher.pdf [accessed 2018-01-16] [WebCite Cache ID 6wW1h8h3Q]

30. Ward DJ, Furber C, Tierney S, Swallow V. Using Framework Analysis in nursing research: a worked example. J Adv Nurs 2013 Nov;69(11):2423-2431. [doi: 10.1111/jan.12127] [Medline: 23517523]
31. Ritchie J, Lewis J, Nicholls CR. Qualitative research practice: A guide for social science students and researchers. In: 2013age. London: Sage; 2013.
32. Swallow V, Newton JLC. How to manage and display qualitative data using 'Framework' and Microsoft® Excel. Journal of Clinical Nursing 2003;12(4):610-612. [doi: 10.1046/j.1365-2702.2003.00728.x]
33. Cai RA, Beste D, Chaplin H, Varakliotis S, Suffield L, Josephs F, et al. Developing and Evaluating JIApp: Acceptability and Usability of a Smartphone App System to Improve Self-Management in Young People With Juvenile Idiopathic Arthritis. JMIR Mhealth Uhealth 2017 Aug 15;5(8):e121 [FREE Full text] [doi: 10.2196/mhealth.7229] [Medline: 28811270]
34. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016 May;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
35. Wyatt JJ. What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist 2015;Royal College of Physicians [FREE Full text]
36. Kenny R, Dooley B, Fitzgerald A. Developing mental health mobile apps: Exploring adolescents' perspectives. Health Informatics J 2016 Jun;22(2):265-275. [doi: 10.1177/1460458214555041] [Medline: 25385165]
37. Liptzin DS. Evolution of Asthma Self-Management Programs in adolescents: From the Crisis Plan to Facebook. The Journal of Paediatrics 2016:179-123.
38. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
39. Miller A, Cafazzo J, Seto E. Gamification design principles in mHealth applications for chronic disease management. Health Informatics Journal 2014:1-10. [doi: 10.1177/1460458214537511]
40. Sardi L, Idri A, Fernández-Alemán J. A Systematic Review of Gamification in e-Health. Journal of Biomedical Informatics; (71). Doi.org/10.1016/j.jbi.05 2017:011-048. [doi: 10.1016/j.jbi.2017.05.011]
41. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
42. Weiser E. Gender Differences in Internet Use Patterns and Internet Application Preferences: A Two-Sample Comparison. CyberPsychology & Behavior. CyberPsychology & Behavior 2004;3(2) [FREE Full text]

Abbreviations

HCP: health care professional
LTC: long-term condition
NRAS: National Rheumatoid Arthritis Society

Edited by G Eysenbach; submitted 14.10.17; peer-reviewed by J McDonagh, A Hall, X Huang, A Rashid; comments to author 03.11.17; accepted 17.11.17; published 19.01.18.

Please cite as:
Waite-Jones JM, Majeed-Ariss R, Smith J, Stones SR, Van Rooyen V, Swallow V
Young People's, Parents', and Professionals' Views on Required Components of Mobile Apps to Support Self-Management of Juvenile Arthritis: Qualitative Study
JMIR Mhealth Uhealth 2018;6(1):e25
URL: http://mhealth.jmir.org/2018/1/e25/
doi:10.2196/mhealth.9179
PMID:29351898

Original Paper

Importance of Active Participation in Obesity Management Through Mobile Health Care Programs: Substudy of a Randomized Controlled Trial

Bumjo Oh[1*], MPH, MD; Ga-Hye Yi[1*], MD; Min Kyu Han[2], MPH, MD; Jong Seung Kim[1], MPH, MD, PhD; Chang Hee Lee[3], PhD; Belong Cho[4], MD, PhD; Hee Cheol Kang[5], MD, PhD

[1]Department of Family Medicine, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul National University College of Medicine, Seoul, Republic Of Korea

[2]SK International Medical Center, Fenghuang Hospital, Wuxi, China

[3]Future IT R&D Laboratory, LG Electronics, Seoul, Republic Of Korea

[4]Department of Family Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic Of Korea

[5]Department of Family Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic Of Korea

[*]these authors contributed equally

Corresponding Author:
Jong Seung Kim, MPH, MD, PhD
Department of Family Medicine
Seoul Metropolitan Government - Seoul National University Boramae Medical Center
Seoul National University College of Medicine
20 Boramae-ro 5 gil, Dongjak-gu
Seoul, 07061
Republic Of Korea
Phone: 82 2 870 2681
Fax: 82 2 831 0714
Email: atenae68@nate.com

Abstract

Background: Due to the prevalence of the westernized dietary pattern and lack of physical activity, the numbers of overweight or obese individuals are increasing, resulting in a growing health burden because of various related diseases. A lifestyle modification approach has additional advantages compared with pharmacological therapies or bariatric surgery. In our randomized controlled trial conducted in 2015, we successfully used a ubiquitous health care (SmartCare) service for patients with metabolic syndrome to achieve a significant weight loss effect. Various useful apps have been developed for the SmartCare Service, which involves using a mobile phone to manage chronic diseases, minimizing time and space restrictions. Many studies have demonstrated weight loss effects using a SmartCare service, but limited data are available regarding the effect of active participation in relation to weight loss.

Objective: We aimed to assess the weight loss effect achieved after using the SmartCare service in terms of adherence and participation. We divided the intervention group of the previous study according to participation level, and analyzed whether there was a significant difference in the outcome.

Methods: We classified participants into 3 groups according to their adherence. Within the intervention group using the SmartCare service, the active group comprised those transmitting anthropometric measurement data using a mobile phone 3 or more times per week or who had a health consultation 5 or more times during a 24-week period. The passive group comprised those who did not adhere to these levels of engagement. The control group comprised those who did not use the SmartCare service. We compared changes in body weight, body mass index (BMI), body fat percentage, waist circumference, and lipid profile among the 3 groups.

Results: We identified 422 participants and analyzed 405, excluding 17 who were missing necessary data for analysis. The active group consisted of 116 participants, compared with 80 in the passive group and 209 in the control group (without SmartCare service). There was a statistically significant difference in improvements to body weight, BMI, body fat percentage, and waist circumference among active participants compared with less active participants and the control group (P<.05). However, the lipid profile changes did not differ significantly between groups.

Conclusions: The level of participation may be related to improved weight-related outcomes, which may improve health outcomes. To maximize the effectiveness of the SmartCare service, encouraging active participation is important.

Trial Registration: Clinicaltrials.gov NCT01344811; https://clinicaltrials.gov/ct2/show/NCT01344811 (Archived by WebCite at http://www.webcitation.org/6alT2MmIB)

(JMIR Mhealth Uhealth 2018;6(1):e2) doi:10.2196/mhealth.8719

KEYWORDS

physical activity; mobile health; metabolic syndrome; self-report; adherence; concordance

Introduction

Being overweight or obese increases the risk of cardiovascular disease, diabetes, cancer, and musculoskeletal disorders, resulting in approximately 3 million deaths worldwide each year [1]. Westernized dietary patterns and a lack of physical activity are considered to be the primary causes. According to the Korea National Health and Nutrition Examination Surveys, the prevalence of obesity in Korean adults was 32.0% in 2011, which was much higher than the 26.9% recorded in 1998 [2]. Therefore, it is important to manage and treat obesity, and it is essential to improve eating and exercise habits before considering medication or surgical treatment options [3]. However, because lifestyle modifications need continuous monitoring and feedback in everyday life, patients can find these modifications difficult and burdensome [4]. An important issue concerns how best to overcome the difficulties in adhering to lifestyle modifications.

Koreans have the highest level of smartphone ownership in the world, and smartphone penetration rate reached 79.5% in 2012 [5]. Recently, the use of a ubiquitous health care (SmartCare) service, equipped with advanced technology for the health care of chronically ill patients with hypertension and diabetes, has been rapidly growing [6]. Numerous studies have revealed that a mobile phone can be used as a strategic tool for weight management in obese patients [7-9]. Unlike desktop computers, a mobile phone provides mobility to users, and also includes sensors such as global positioning systems and acceleration sensors, as well as being applicable to various mobile health care services. It is also possible to access the Internet easily without time and space restrictions, thus enabling real-time one-on-one consultation and communication with health care experts.

Previously, we conducted a randomized controlled trial to evaluate the effect of a SmartCare service on weight management in obese patients with metabolic syndrome (NCT01344811) [10]. Among adults over 20 years of age who visited Seoul National University Hospital or the Severance Hospital, both in Seoul, South Korea, obese patients with metabolic syndrome were identified for possible inclusion in this study. The prevalence of metabolic syndrome in Korean adults continues to increase, which increases cardiovascular morbidity and mortality and all-cause mortality [11-13].

That study was, to our knowledge, the first government-funded project in Korea undertaken to estimate the usefulness of a SmartCare service in managing chronic disease. It concluded that weight control using the SmartCare service was significantly better than for a control group using conventional methods. However, the effect of adherence to the SmartCare service has not been analyzed among participants.

Recent studies related to mobile health care services have focused primarily on comparing information and communication technology intervention efficacy between users and nonusers, so there is little research on whether adherence and weight loss outcomes are positively correlated. The purpose of this study was to clarify the relationship between adherence and weight loss outcomes through further analysis of our previous study.

Prior to the study, we hypothesized that a higher level of participation would produce a better outcome, and we attempted to set criteria for intensive participation. As family medical practitioners involved in patient education, we aimed not only to recommend the SmartCare service, but also to provide effective participation criteria.

Methods

Recruitment

We advertised our study using institutional review board (IRB)-approved banners, posters, and leaflets placed in the hospitals' lobbies.

According to the Adult Treatment Panel III criteria, using waist circumference cutoff modifications for Asian populations as suggested in the Asia-Pacific guidelines, we defined metabolic syndrome as having 3 or more of the following factors [14]: (1) central obesity: waist circumference 90 cm or more in men and 80 cm or more in women; (2) hypertriglyceridemia: triglyceride 8.325 mmol/L or more; (3) high-density lipoprotein (HDL) cholesterol less than 2.22 mmol/L in men and less than 2.775 mmol/L in women; (4) hypertension: blood pressure 130/85 mmHg or more, or taking antihypertensive medication; and (5) hyperglycemia: fasting glucose 5.55 mmol/L or more, or taking antidiabetic medication.

The World Health Organization Western Pacific Region recommends defining obesity in Asian populations as those individuals with a body mass index (BMI) 25 kg/m^2 or greater [15]. The Korean Society for the Study of Obesity also investigated BMI cutoff figures for obesity-related disease and adopted the World Health Organization's definition. Nowadays, Korean government organizations officially use this definition when defining and implementing health policies regarding obesity in Korea [16].

However, we excluded from this study individuals taking thyroid hormone or antiobesity medication, which can affect weight,

along with patients with insulin-dependent diabetes with liver functional abnormality (with a liver somatic index over 3 times higher than the normal maximum level) or with renal functional impairment (with a creatinine level over one and a half times higher than the normal maximum level), pregnant women, and inpatients.

We assigned eligible participants to 2 groups, with equal probability, according to a randomization code. The randomization code was prepared using a block randomization method, stratified by a statistician in a clinical trial center (C&R Research, Seoul, South Korea). As this study was an open-label trial, blinding was not undertaken.

The IRB of Seoul National University Hospital approved this study (IRB number: h-1009-095-333).

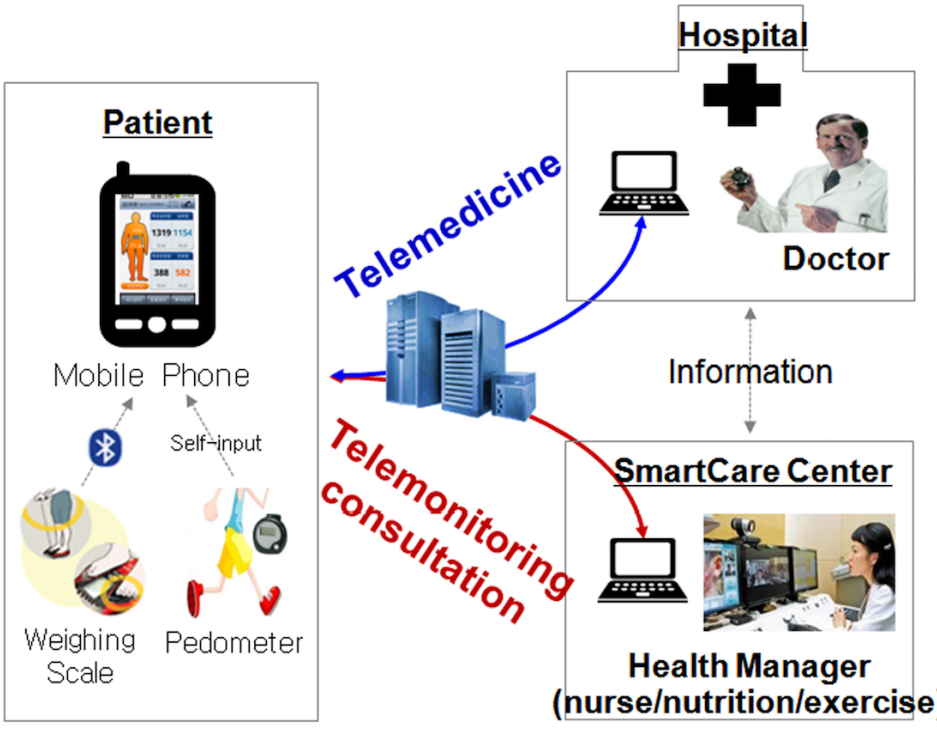
Intervention Using a SmartCare Service

We randomly allocated participants into 2 groups. The intervention group was provided with a mobile phone equipped with a SmartCare app, a body composition meter (InBody IH-U070B, InBody Co Ltd, Seoul, South Korea) with Bluetooth function, and a pedometer (MP-100, Yamasa Co, Ltd, Okayama, Japan).

Each participant measured his or her own body weight and body composition (skeletal muscle mass, fat mass) using the provided body composition monitor at the same hour every day, if possible (or at least 3 times a week), after waking up and before breakfast. After measuring the relevant values with the body composition monitor, the participants placed the transmission terminal (Bluetooth) of the remote monitoring device near the transmission terminal of the mobile phone to transmit the measurement data to the central server in the SmartCare center via mobile phone (wireless network). Each participant carried a pedometer on the waist from the time they woke until the time they went to bed. Participants checked their activity level, indicated as the number of steps taken, at the same time every day, if possible, and entered this into the mobile phone (data entry before going to bed was recommended). The entered data were then automatically transmitted to the central server in the SmartCare Center. Physicians or health care personnel at the SmartCare Center could retrieve hospital admission information, treatment records, diagnosed diseases identification, diagnostic examinations and functional test results, and prescription information of the test participants through connecting to the hospital information system, with participant consent. The central server in the SmartCare Center transmitted the feedback, based on the measured body weight and body composition, to the mobile phones of the participants, according to the algorithm of the clinical decision support system. The participants were able to immediately check the interpretations and recommendations, based on their measured values, through their mobile phones.

Trained consultants (nurses, exercise specialists, and clinical dietitians) at the SmartCare Center undertook various health consultations through the patients' telephone inquiries concerning disease management, health education, recommended exercise, medication, and appropriate nutrition. Monthly and weekly health reports, based on the individual patient's measured values and daily living records, were sent directly to the patients through the SmartCare system. Additionally, in instances where medical treatment was received at Seoul National University Hospital or the Severance Hospital, the physician providing the medical care had access to the information (Figure 1).

Figure 1. Model of SmartCare service.



Comparison Group

We provided the comparison group only with anthropometric instruments and pedometers.

We provided body weight scales and pedometers to the participants assigned to the control group. Body weight journals were distributed to the participants, and each participant self-measured and recorded his or her daily weight and waist size (a minimum of 3 times per week) at the same time (before breakfast). They also wore a pedometer during daily activities, which started from the time they woke in the morning until bedtime. They were instructed to check and record their daily walking distances on the record sheet just before going to bed. Additionally, the participants in the control group had the same hospital visit schedule as that of the intervention group at 12- and 24-week intervals, and received anthropometry, consultations with physicians, and information about their nutrition and exercise.

As with the intervention group, individual health care counseling services were provided, but only at the hospital.

As an incentive for registering as study participants for both the intervention and comparison groups, the Korean Ministry of Commerce, Industry and Energy national project budget covered all expenses for medical treatment, medication, transportation, and communication (mobile phone provision and use).

Study Design

We requested the participants to visit the hospitals 4 times during the 24-week period. Except when screening was performed, body weight, body composition, and blood pressure measurements were recorded. A hematology test was performed, and changes in their life habits (eg, diet intake and physical activity) were recorded 3 times during the test period: once on the date of random allocation, once in week 12, and once in week 24 (Figure 2).

Analysis consisted of 2 group sets: intention-to-treat (ITT) and per protocol. The ITT set included all participants who were enrolled in this clinical trial and were randomly allocated. When analyzing efficacy, we included participants in the treatment group to which they had been randomly allocated, regardless of the actual treatment they received. Among the participants in the ITT set, we included those who completed this clinical trial without a material breach of the protocol in the per protocol set (Figure 3).

Additional Analysis According to Medical Adherence

We classified the intervention group into an active participation group and a passive participation group, according to adherence. The definition of adherence was divided into 2 parts as follows: (1) measurement of anthropometric data and input of pedometer data directly into the mobile phone 3 or more times a week, and (2) taking part in a health consultation 5 or more times during a 24-week period.

Figure 2. Flowchart of the intervention.

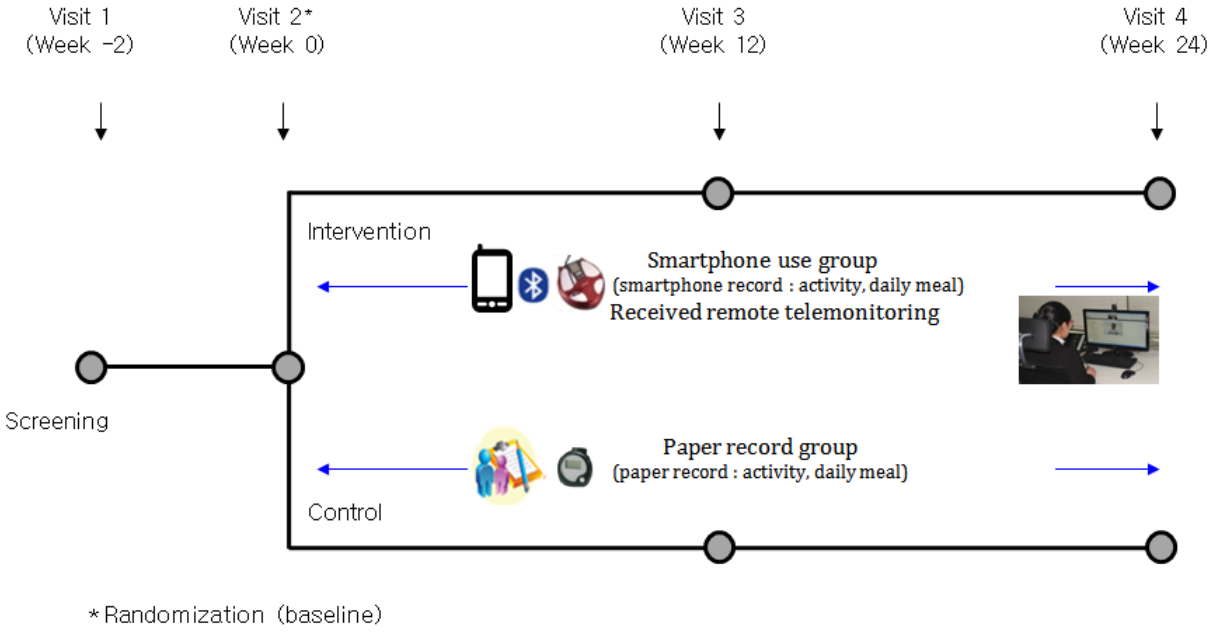
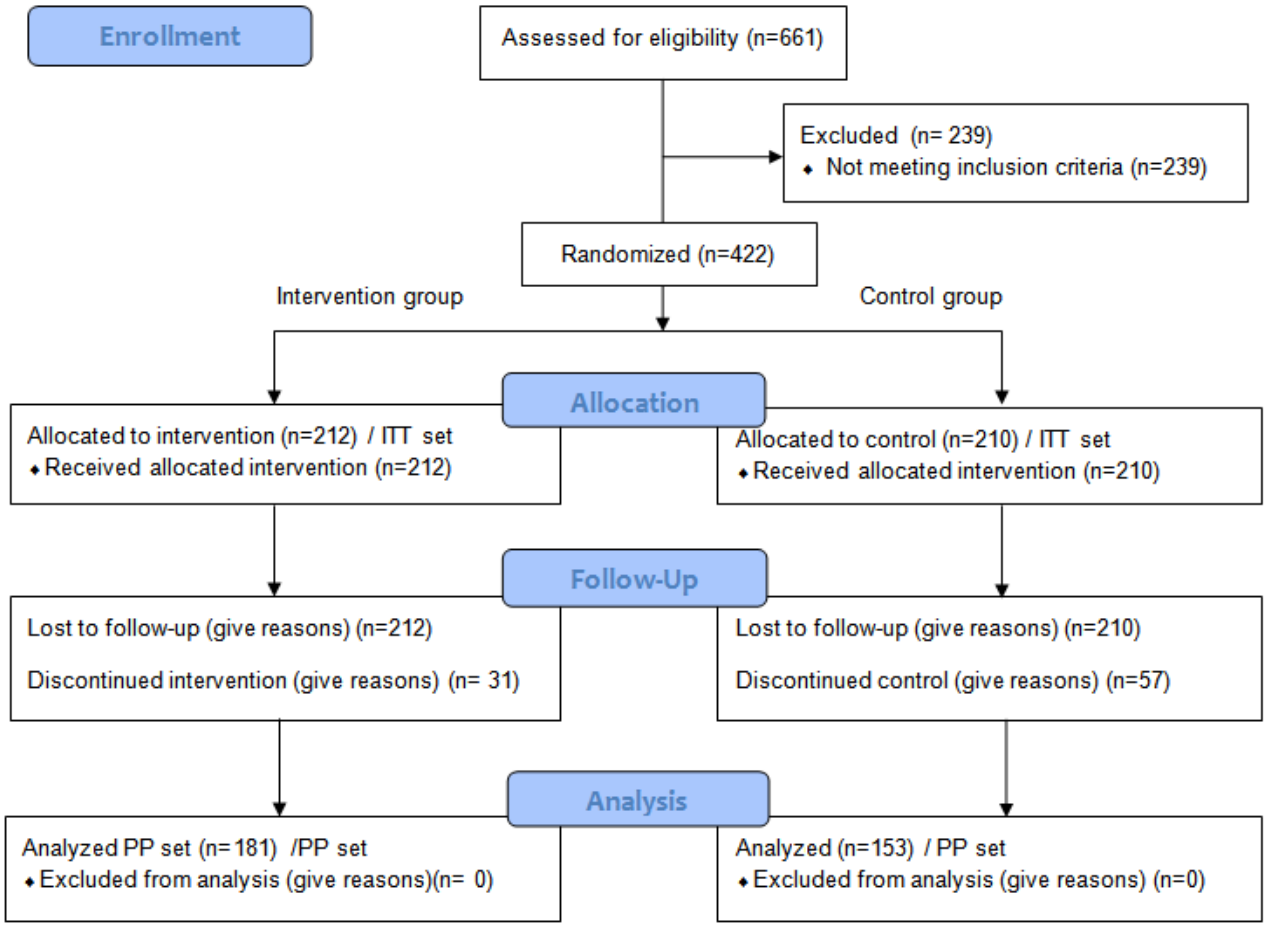


Figure 3. Selection of the study participants. ITT: intention-to-treat; PP: per protocol.



Measurement Variables

At the time of screening, demographic information (age, sex, smoking, alcohol consumption, and other relevant information), medical history, and medication history of the participants were investigated and recorded. Additionally, an electrocardiogram was recorded at the screening visit after the participant had been resting for at least 5 minutes. When test results were clinically significant, the investigator determined whether to enroll the participant in the study.

Laboratory tests were conducted at screening, baseline, week 12, and week 24 for alanine aminotransferase, aspartate aminotransferase, creatinine, lipid profile (total cholesterol, HDL cholesterol, and triglyceride), fasting glucose, and hemoglobin A_{1c}. However, alanine aminotransferase, aspartate aminotransferase, and creatinine tests were performed only at screening to determine trial eligibility. Lipid profiles and blood glucose tests were performed after the participant had been fasting.

We evaluated weight change, the primary outcome, using body fat percentage at baseline and at 12 and 24 weeks, as measured by nurses using portable bioelectrical impedance analysis (InBody U20, InBody Co Ltd).

We assessed and categorized the level of physical activity using the International Physical Activity Questionnaire at baseline, 12 weeks, and 24 weeks. The amount of physical activity each week was calculated with a continuous variable, namely the metabolic equivalent task (MET), as follows: total METs min/week = [walking METs × min × days] + [moderate METs × min × days] + [vigorous METs × min × days]) [17].

For measuring caloric intake variables, daily meal record cards (3-day recall dietary assessments) were distributed to the participants during their initial visit, and the participants were instructed to write their own 3-day meal record just before the baseline visit, the next visit after 12 weeks, and the final visit at 24 weeks. The self-completed daily meal records were collected from the participants during their final visit, and dietitians calculated caloric intake using the nutrient evaluation program CAN-Pro 3.0 (The Korean Nutrition Society) [18].

Statistical Analysis

The effectiveness analysis comprised the ITT population, which included all randomly allocated participants. We imputed missing data using the last observation carried forward method.

We calculated descriptive statistics (number of observed participants, mean [standard deviation (SD)], and median [range]) on body weight measured at baseline and at the end point (24 weeks), and the changes in measured values at week 24 compared with baseline for each group. To identify the difference between the groups with respect to body weight changes at week 24 compared with baseline, we used analysis

of covariance (ANCOVA), with the clinical trial institution that recruited participants and body weight at baseline as covariates.

For continuous data (changes in BMI, body fat percentage, waist measurement, lipid profile, blood pressure, number of metabolic syndrome elements, diet intake in kilocalories, physical activity, number of steps taken, and weight to measure physical activity), we determined descriptive statistics (number of observed participants, mean [SD], and median range]) per group. To identify the difference between the groups in the changes at week 24 compared with baseline, we performed ANCOVA or rank transformation ANCOVA. When conducting ANCOVA, we set the clinical trial institution that recruited participants and the baseline values of relevant parameters as covariates.

All analyses were conducted using SAS version 9.3 (SAS Institute). All statistical tests were performed at a 2-sided significance level of .05.

Results

Baseline Characteristics

We identified 422 patients able to use mobile phones, and who were willing to participate in the study and sign informed consent forms, for recruitment. All participants received the SmartCare service, and we excluded from the final analysis 17 participants who did not enter any data for 24 weeks. Following the definition of adherence already described, we used 2 criteria for dividing the active and passive participation groups. First, we analyzed the demographic characteristics according to the number of weekly mean anthropometric measurements. Second, we analyzed the data according to total counseling frequency, age, sex, BMI, alcohol consumption, education level, and smoking status did.

Efficacy Evaluation Summary

Changes in Anthropometric Data During the 24-Week Period

First, we examined differences in characteristics according to the weekly mean number of anthropometric measurements (Figure 4).

Following classification according to self-reporting frequency, there were 116 active participants and 80 less active participants (Table 1). The average weight loss effect during the whole period was greatest in the active participation group, lower in the control group, and lowest in the less active participation group. There was no statistically significant difference between the control group and the less active participation group (Table 1). The mean change of body weight was –3.18 (SD 0.29) kg in the active group, –0.70 (SD 0.35) kg in the less active group, and –0.82 (SD 0.23) kg in the control group. BMI, body fat percentage, and waist circumference were most enhanced for those in the active participation group, less enhanced in the less active participation group, and least enhanced in the control group. In particular, there was no difference in body weight and BMI at baseline between the active participant group and the control group, but there was a statistically significant difference in body weight, BMI, body fat percentage, and waist circumference at the end of 24 weeks (Table 1).

Figure 4. Changes in anthropometric data during the 24-week period analyzed according to weekly mean number of anthropometric measurements. Error bars indicate standard deviation. BMI: body mass index.

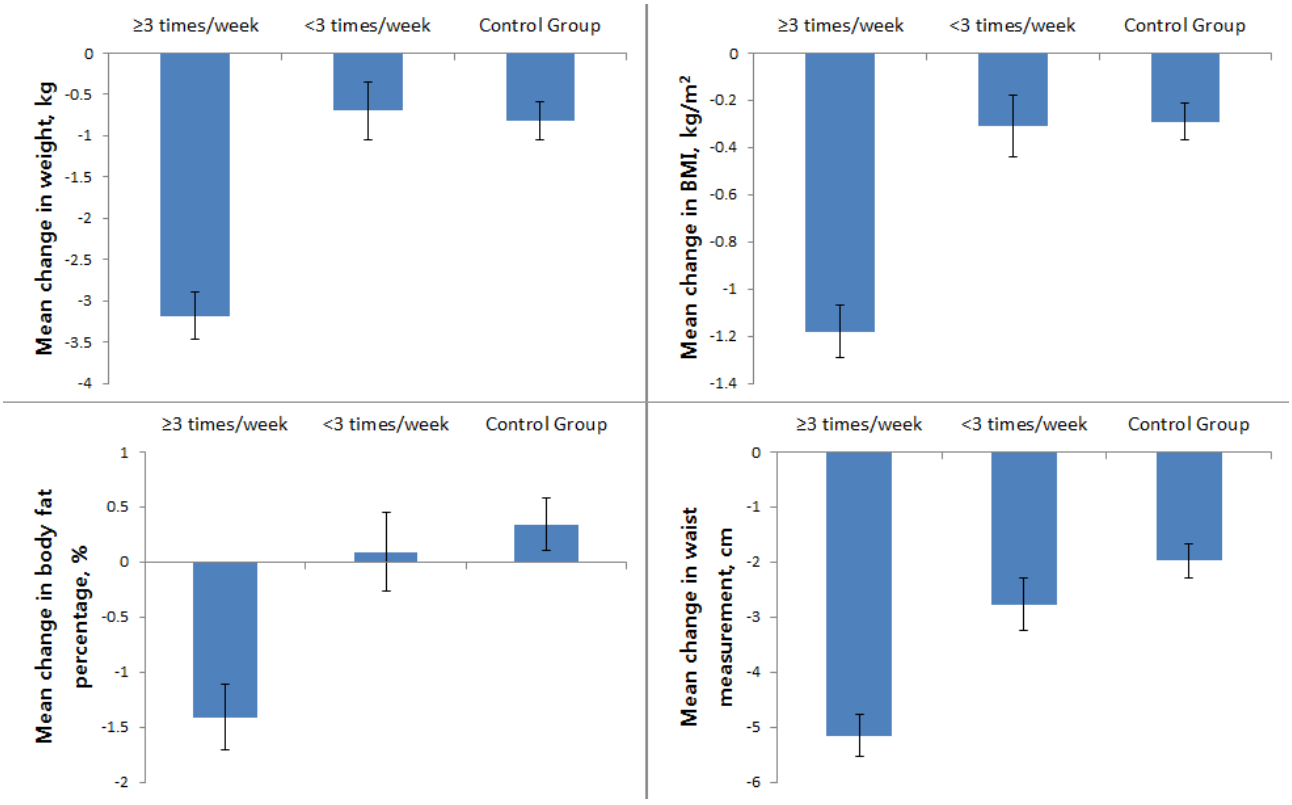


Table 1. Demographic and other pretreatment characteristics (intention-to-treat set) by the weekly mean number of anthropometric measurements (n=405).

Demographic and other pretreatment characteristics	Intervention group		Control group (C)	Between-group *P* value		
	≥3 times a week (A)	<3 times a week (B)		A-B	A-C	B-C
Age (years)						
Total, n (%)	116 (28.6)	80 (20.0)	209 (51.6)			
Mean (SD[a])	50.45 (12.24)	42.11 (12.33)	50.36 (14.27)	<.001[b]	.96[b]	<.001[b]
Median	52.50	41.50	52.00			
Range	23.00-71.00	20.00-72.00	21.00-82.00			
Range, n (%)						
20-29	5 (4.3)	14 (17.5)	18 (8.6)	<.001[c]	.55[c]	.001[c]
30-39	21 (18.1)	23 (28.8)	40 (19.1)			
40-49	21 (18.1)	20 (25.0)	33 (15.8)			
50-59	40 (34.5)	15 (18.8)	61 (29.2)			
≥60	29 (25.0)	8 (10.0)	57 (27.3)			
Sex, n (%)						
Total	116 (28.6)	80 (20.0)	209 (51.6)			
Male	57 (49.1)	44 (55.0)	101 (48.3)	.42[c]	.89[c]	.31[c]
Female	59 (50.9)	36 (45.0)	108 (51.7)			
Body mass index (kg/m^2)						
Total, n (%)	116 (28.6)	80 (20.0)	209 (51.6)			
Mean (SD)	28.77 (3.09)	29.72 (3.11)	29.40 (3.39)	.03[b]	.1[b]	.45[b]
Median	28.00	29.20	28.90			
Range	24.90-41.00	25.00-41.40	24.90-41.80			
Body weight (kg)						
Total, n (%)	116 (28.6)	80 (20.0)	209 (51.6)			
Mean (SD)	77.48 (12.30)	83.86 (14.51)	79.74 (15.28)	.001[b]	.15[b]	.04[b]
Median	76.60	83.20	76.90			
Range	55.20-128.50	58.60-135.40	54.00-141.10			
Height (cm)						
Total, n (%)	116 (28.6)	80 (20.0)	209 (51.6)			
Mean (SD)	163.82 (8.73)	167.49 (9.29)	164.12 (10.61)	.005[b]	.78[b]	.01[b]
Median	165.00	169.20	163.00			
Range	146.00-182.00	144.50-188.90	142.80-189.00			
Smoking status, n (%)						
Total	115 (28.4)	80 (20.0)	203 (51.6)			
Nonsmoker	72 (62.6)	47 (58.8)	129 (63.6)	.02[c]	.02[c]	.65[c]
Former smoker	34 (29.6)	16 (20.0)	40 (19.7)			
Smoker	9 (7.8)	17 (21.3)	34 (16.8)			
Drinking status, n (%)						
Total	115 (28.4)	80 (20.0)	206 (51.6)			
Nondrinker	51 (44.4)	20 (25.0)	84 (40.8)	.01[c]	.68[c]	.03[c]
Former drinker	10 (8.7)	5 (6.3)	15 (7.3)			

Demographic and other pretreatment characteristics	Intervention group		Control group (C)	Between-group *P* value		
	≥3 times a week (A)	<3 times a week (B)		A-B	A-C	B-C
Drinker	54 (47.0)	55 (68.8)	107 (51.9)			
Completed education level, n (%)						
Total	116 (28.6)	80 (20.0)	209 (51.6)			
Elementary school	4 (3.5)	0	21 (10.2)	.11[d]	.15[c]	.001[d]
Middle school	6 (5.2)	3 (3.85)	15 (7.2)			
High school	39 (33.6)	19 (23.8)	65 (31.1)			
University	67 (57.8)	58 (72.5)	108 (51.7)			

[a]SD: standard deviation

[b]2-sample *t* test.

[c]Pearson chi-square test.

[d]Fisher exact test.

Second, we examined differences in characteristics according to the total number of health consultations (Figure 5).

Following classification according to the total number of health consultations, there were 98 active participants and 98 less active participants (Table 2). Weight, BMI, body fat percentage, and waist circumference were highest in the active participation group, lower in the less active participation group, and lowest in the control group. The mean change of body weight was –2.81 (SD 0.32 kg in the active group, –1.54 (SD 0.32) kg in the less active group, and –0.81 (SD 0.24) kg in the control group. In addition, improvements in body weight, BMI, body fat percentage, and waist circumference differed significantly between the active and the less active participation groups. Even where participants received the same SmartCare service, it was clear that the outcome depended on the participation level (Table 2).

Changes in Biomarker (Lipid Panel) During the 24-Week Period

When we analyzed biomarkers according to the weekly mean number of anthropometric measurements and the total number of health consultations, we found no statistically significant differences between groups (data not shown).

Figure 5. Changes in anthropometric data during the 24-week period analyzed according to total number of health consultations. Error bars indicate standard deviation. BMI: body mass index.

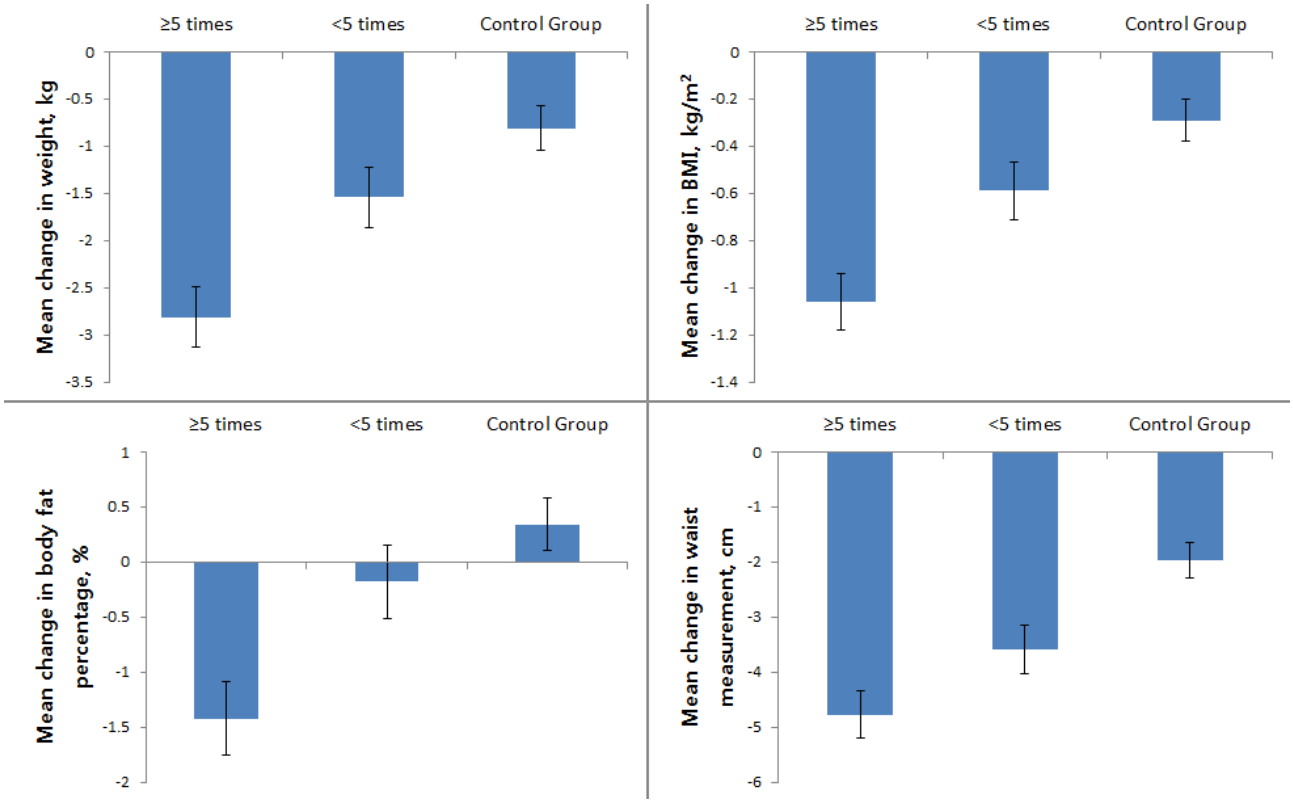


Table 2. Demographic and other pretreatment characteristics (intention-to-treat set) by total number of health consultations (n=405).

Demographic and other pretreatment characteristics	Intervention group		Control group (C)	Between-group *P* value		
	≥5 times in total period (A)	<5 times in total period (B)		A-B	A-C	B-C
Age (years)						
Total, n (%)	98 (24.2)	98 (24.2)	209 (51.6)			
Mean (SD[a])	49.45 (12.29)	44.64 (13.14)	50.36 (14.27)	.009[b]	.59[b]	.001[b]
Median	50.50	44.00	52.00			
Range	23.00-72.00	20.00-71.00	21.00-82.00			
Range, n (%)						
20-29	6 (6.1)	13 (13.3)	18 (8.6)	.17[c]	.67[c]	.07[c]
30-39	19 (19.4)	25 (25.5)	40 (19.1)			
40-49	19 (19.4)	22 (22.5)	33 (15.8)			
50-59	33 (33.7)	22 (22.5)	61 (29.2)			
≥60	21 (21.4)	16 (16.3)	57 (27.3)			
Sex, n (%)						
Total	98 (24.2)	98 (24.2)	209 (51.6)			
Male	53 (54.1)	48 (49.0)	101 (48.3)	.48[c]	.35[c]	.91[c]
Female	45 (45.9)	50 (51.0)	108 (51.7)			
Body mass index (kg/m^2)						
Total, n (%)	98 (24.2)	98 (24.2)	209 (51.6)			
Mean (SD)	28.82 (3.30)	29.50 (2.93)	29.40 (3.39)	.13[b]	.16[b]	.8[b]
Median	28.00	29.00	28.90			
Range	25.00-41.40	24.90-37.90	24.90-41.80			
Body weight (kg)						
Total, n (%)	98 (24.2)	98 (24.2)	209 (51.6)			
Mean (SD)	79.00 (13.83)	81.17 (13.30)	79.74 (15.28)	.26[b]	.68[b]	.46[b]
Median	77.25	80.95	76.90			
Range	55.20-128.50	55.50-135.40	54.00-141.10			
Height (cm)						
Total, n (%)	98 (24.2)	98 (24.2)	209 (51.6)			
Mean (SD)	165.17 (8.79)	165.47 (9.48)	164.12 (10.61)	.82[b]	.36[b]	.28[b]
Median	167.00	165.35	163.00			
Range	144.50-181.00	148.00-188.90	142.80-189.00			
Smoking status, n (%)						
Total	98 (24.2)	97 (24.2)	203 (51.6)			
Nonsmoker	57 (58.2)	62 (63.9)	129 (63.6)	.01[c]	.01[c]	.87[c]
Former smoker	33 (33.7)	17 (17.5)	40 (19.7)			
Smoker	8 (8.2)	18 (18.6)	34 (16.8)			
Drinking status, n (%)						
Total	98 (24.2)	97 (24.2)	206 (51.6)			
Nondrinker	33 (33.7)	38 (39.2)	84 (40.8)	.36[c]	.41[c]	.72[c]
Former drinker	10 (10.2)	5 (5.2)	15 (7.3)			

Demographic and other pretreatment characteristics	Intervention group		Control group (C)	Between-group *P* value		
	≥5 times in total period (A)	<5 times in total period (B)		A-B	A-C	B-C
Drinker	55 (56.1)	54 (55.7)	107 (51.9)			
Completed education level, n (%)						
Total	98 (24.2)	98 (24.2)	209 (51.6)			
Elementary school	1 (1.0)	3 (3.1)	21 (10.1)	.37[d]	.02[c]	.04[d]
Middle school	4 (4.1)	5 (5.1)	15 (7.2)			
High school	34 (34.7)	24 (24.5)	65 (31.1)			
University	59 (60.2)	66 (67.4)	108 (51.7)			

[a]SD: standard deviation

[b]2-sample *t* test.

[c]Pearson chi-square test.

[d]Fisher exact test.

Discussion

Principal Findings

One previous study found that self-monitoring of body weight in the workplace had a preventive effect on weight gain [19]. That study concluded that just increasing the self-measurement frequency of body weight had a preventive effect, that is, weight management improved through measurement adherence. For this reason, we wanted to identify the cutoff value of adherence, given that people with better adherence during the 24-week study period had a better outcome. As a measure of participation, we identified the weekly mean number of anthropometric measurements and the total number of health consultations. As we had hypothesized, the higher the participation, the greater the improvement in anthropometric measurements.

The high frequency of physical measurement and consultation frequency indicates that there was a strong willingness among participants to manage their obesity.

Health care information technology (IT), also referred to as medical IT, encompasses the process of storing, analyzing, and delivering information, data, and knowledge from all activities related to health care, using information processing technology and networks. Demand is rapidly increasing in line with developments in areas of medical care, such as population aging, early diagnosis and treatment of diseases, and preventive medicine. The medical IT convergence technologies that are being actively developed in Korea include home care, using a blood pressure monitor and blood glucose meter; home health care, using a mobile phone to measure and transmit biometric information to a service center; and wearable health care, involving a wearable biosignal measurement system [20].

Many studies have already demonstrated the utility of managing chronic diseases and obesity through health care programs that use IT [21-26]. However, because programs are implemented outside the hospital, continuous monitoring is needed to encourage active participation, and it is important to inform patients that the higher the participation, the better the health improvement.

Comparison With Prior Research

Many studies have shown that close adherence to medical treatment, or to lifestyle modification, results in a favorable outcome [27-29]. Mobile phone health care is a new strategic tool for chronic disease management, and numerous studies have demonstrated the effectiveness of mobile phone health care. However, our hypothesis was that active participation would maximize the effects of mobile phone health care, about which there has been little previous research. In our study, we defined adherence as measuring anthropometric data 3 or more times a week, or taking part in voluntary health consultations at least 5 times in 24 weeks. However, the definition of adherence can vary [30-32]. Further studies should be conducted to demonstrate the relevance of adherence and the effectiveness of mobile health care.

Limitations and Strengths

Concerning the lipid profile, although participation in the SmartCare service was high, the improvement effect was not significant. Previous studies have attempted to improve patients' lipid profile through lifestyle modification, and these studies have usually involved following the changing pattern of biomarkers from 3 months to 5 years [33,34]. It is possible that the 24-week period set in this study was not sufficient to show changes in lipid profiles. Another plausible reason for the reduced improvement in lipid profiles is that the intervention to improve dietary habits was not appropriately targeted. A more active intervention in respect of eating habits could also improve biomarkers, including lipid profiles.

Although there was no statistically significant difference, all total cholesterol, low-density lipoprotein cholesterol, HDL cholesterol, and triglyceride levels tended to decrease. We found that changes in HDL cholesterol and triglyceride levels, both of which are key factors of metabolic syndrome, were close to a *P* value of .05, although not significant, demonstrating that the SmartCare service may be able to help patients with metabolic syndrome make effective lifestyle modifications.

Conclusions

Through initial and subsequent studies, we conclude that using the SmartCare service is an effective way to control the weight of obese patients with metabolic syndrome. In addition, the improvement effect depended on adherence, defined as entering anthropometric information 3 or more times a week or participating in voluntary health consultations at least 5 times in 24 weeks.

Acknowledgments

This study was supported with a research grant from the Ministry of Trade, Industry and Energy of South Korea in 2010 (1003518).

Conflicts of Interest

None declared.

References

1. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Body Mass Index). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. Lancet 2011 Feb 12;377(9765):557-567. [doi: 10.1016/S0140-6736(10)62037-5] [Medline: 21295846]
2. Kim CS, Ko S, Kwon H, Kim NH, Kim JH, Lim S, Taskforce Team of Diabetes Fact Sheet of the Korean Diabetes Association. Prevalence, awareness, and management of obesity in Korea: data from the Korea national health and nutrition examination survey (1998-2011). Diabetes Metab J 2014 Feb;38(1):35-43 [FREE Full text] [doi: 10.4093/dmj.2014.38.1.35] [Medline: 24627826]
3. Lang A, Froelicher ES. Management of overweight and obesity in adults: behavioral intervention for long-term weight loss and maintenance. Eur J Cardiovasc Nurs 2006 Jun;5(2):102-114. [doi: 10.1016/j.ejcnurse.2005.11.002] [Medline: 16406709]
4. Klein S, Sheard NF, Pi-Sunyer X, Daly A, Wylie-Rosett J, Kulkarni K, American Diabetes Association, North American Association for the Study of Obesity, American Society for Clinical Nutrition. Weight management through lifestyle modification for the prevention and management of type 2 diabetes: rationale and strategies: a statement of the American Diabetes Association, the North American Association for the Study of Obesity, and the American Society for Clinical Nutrition. Diabetes Care 2004 Aug;27(8):2067-2073. [Medline: 15277443]
5. Yonhap News. S. Korea tops smartphone penetration rate in 2012 Internet. Seoul, South Korea: The Korea Herald; 2013. URL: http://www.koreaherald.com/view.php?ud=20130625000550 [accessed 2017-12-20] [WebCite Cache ID 6vqaveXSE]
6. Kratzke C, Cox C. Smartphone technology and apps: rapidly changing health promotion. Glob J Health Educ Promot 2012;15(1):72.
7. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
8. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
9. Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. J Telemed Telecare 2014 Sep;20(6):339-349. [doi: 10.1177/1357633X14537777] [Medline: 24875928]
10. Oh B, Cho B, Han MK, Choi H, Lee MN, Kang H, et al. The effectiveness of mobile phone-based care for weight control in metabolic syndrome patients: randomized controlled trial. JMIR Mhealth Uhealth 2015;3(3):e83 [FREE Full text] [doi: 10.2196/mhealth.4222] [Medline: 26293568]
11. Isomaa B, Almgren P, Tuomi T, Forsén B, Lahti K, Nissén M, et al. Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care 2001 Apr;24(4):683-689. [Medline: 11315831]
12. Haffner SM, Valdez RA, Hazuda HP, Mitchell BD, Morales PA, Stern MP. Prospective analysis of the insulin-resistance syndrome (syndrome X). Diabetes 1992 Jun;41(6):715-722. [Medline: 1587398]
13. Lim S, Shin H, Song JH, Kwak SH, Kang SM, Won YJ, et al. Increasing prevalence of metabolic syndrome in Korea: the Korean National Health and Nutrition Examination Survey for 1998-2007. Diabetes Care 2011 Jun;34(6):1323-1328 [FREE Full text] [doi: 10.2337/dc10-2109] [Medline: 21505206]
14. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA 2001 May 16;285(19):2486-2497. [Medline: 11368702]
15. World Health Organization, Regional Office for the Western Pacific. The Asia-Pacific Perspective: Redefining Obesity and its Treatment. Sydney, Australia: Health Communications Australia; 2000. URL: http://iris.wpro.who.int/bitstream/handle/10665.1/5379/0957708211_eng.pdf [WebCite Cache ID 6vhsmsFPd]

16. Oh SW, Shin S, Yun YH, Yoo T, Huh B. Cut-off point of BMI and obesity-related comorbidities and mortality in middle-aged Koreans. Obes Res 2004 Dec;12(12):2031-2040. [doi: 10.1038/oby.2004.254] [Medline: 15687405]
17. Chun MY. Validity and reliability of Korean version of International Physical Activity Questionnaire short form in the elderly. Korean J Fam Med 2012 May;33(3):144-151 [FREE Full text] [doi: 10.4082/kjfm.2012.33.3.144] [Medline: 22787536]
18. Shim Y, Paik H. Reanalysis of 2007 Korean National Health and Nutrition Examination Survey (2007 KNHANES) results by CAN-Pro 3 Nutrient Database. Korean J Nutr 2009;42(6):577-595 [FREE Full text] [doi: 10.4163/kjn.2009.42.6.577]
19. VanWormer JJ, Linde JA, Harnack LJ, Stovitz SD, Jeffery RW. Self-weighing frequency is associated with weight gain prevention over 2 years among working adults. Int J Behav Med 2012 Sep;19(3):351-358 [FREE Full text] [doi: 10.1007/s12529-011-9178-1] [Medline: 21732212]
20. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR Mhealth Uhealth 2017 Jul 20;5(7):e94 [FREE Full text] [doi: 10.2196/mhealth.7088] [Medline: 28729235]
21. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
22. Choi E, Jeong Y. U-health for management of chronic diseases-physical activity and therapeutic exercise. J Korean Med Assoc 2009;52(12):1154-1163 [FREE Full text] [doi: 10.5124/jkma.2009.52.12.1154]
23. Song T, Ryu S, Lee SH. U-health service for managing chronic disease: a case study on managing metabolic syndrome in a health center in South Korea. Healthc Inform Res 2011 Dec;17(4):260-266 [FREE Full text] [doi: 10.4258/hir.2011.17.4.260] [Medline: 22259728]
24. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
25. Park DK, Kim JH, Kim JK, Jung EY, Lee YH. U-health service model for managing health of chronic patients in multi-platform environment. J Korea Contents Assoc 2011;11(8):23-32.
26. Wang J, Wang Y, Wei C, Yao NA, Yuan A, Shan Y, et al. Smartphone interventions for long-term health management of chronic diseases: an integrative review. Telemed J E Health 2014 Jun;20(6):570-583. [doi: 10.1089/tmj.2013.0243] [Medline: 24787747]
27. Jones RCM, Hyland ME, Hanney K, Erwin J. A qualitative study of compliance with medication and lifestyle modification in chronic obstructive pulmonary disease (COPD). Prim Care Respir J 2004 Sep;13(3):149-154 [FREE Full text] [doi: 10.1016/j.pcrj.2004.05.006] [Medline: 16701658]
28. Alm-Roijer C, Stagmo M, Udén G, Erhardt L. Better knowledge improves adherence to lifestyle changes and medication in patients with coronary heart disease. Eur J Cardiovasc Nurs 2004 Dec;3(4):321-330. [doi: 10.1016/j.ejcnurse.2004.05.002] [Medline: 15572021]
29. Sokol MC, McGuigan KA, Verbrugge RR, Epstein RS. Impact of medication adherence on hospitalization risk and healthcare cost. Med Care 2005 Jun;43(6):521-530. [Medline: 15908846]
30. DiMatteo MR, Giordani PJ, Lepper HS, Croghan TW. Patient adherence and medical treatment outcomes: a meta-analysis. Med Care 2002 Sep;40(9):794-811. [doi: 10.1097/01.MLR.0000024612.61915.2D] [Medline: 12218770]
31. Loh A, Leonhart R, Wills CE, Simon D, Härter M. The impact of patient participation on adherence and clinical outcome in primary care of depression. Patient Educ Couns 2007 Jan;65(1):69-78. [doi: 10.1016/j.pec.2006.05.007] [Medline: 17141112]
32. Rasmussen JN, Chong A, Alter DA. Relationship between adherence to evidence-based pharmacotherapy and long-term mortality after acute myocardial infarction. JAMA 2007 Jan 10;297(2):177-186. [doi: 10.1001/jama.297.2.177] [Medline: 17213401]
33. Yamaoka K, Tango T. Effects of lifestyle modification on metabolic syndrome: a systematic review and meta-analysis. BMC Med 2012 Nov 14;10:138 [FREE Full text] [doi: 10.1186/1741-7015-10-138] [Medline: 23151238]
34. Thoma C, Day CP, Trenell MI. Lifestyle interventions for the treatment of non-alcoholic fatty liver disease in adults: a systematic review. J Hepatol 2012 Jan;56(1):255-266 [FREE Full text] [doi: 10.1016/j.jhep.2011.06.010] [Medline: 21723839]

Abbreviations

ANCOVA: analysis of covariance
BMI: body mass index
HDL: high-density lipoprotein
IRB: institutional review board
IT: information technology
ITT: intention-to-treat

MET: metabolic equivalent task
SD: standard deviation

Edited by G Eysenbach; submitted 19.08.17; peer-reviewed by S Boucher, PH Lin; comments to author 14.09.17; revised version received 15.11.17; accepted 30.11.17; published 03.01.18.

Please cite as:
Oh B, Yi GH, Han MK, Kim JS, Lee CH, Cho B, Kang HC
Importance of Active Participation in Obesity Management Through Mobile Health Care Programs: Substudy of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2018;6(1):e2
URL: http://mhealth.jmir.org/2018/1/e2/
doi:10.2196/mhealth.8719
PMID:29298749

Original Paper

Face-to-Face Versus Mobile Versus Blended Weight Loss Program: Randomized Clinical Trial

Emalie Hurkmans[1,2], PhD; Christophe Matthys[3,4,5], PhD; An Bogaerts[6], PhD; Leonie Scheys[4,5], MSc; Karlien Devloo[1], MScPT; Jan Seghers[1], PhD

[1]Department of Movement Sciences, University of Leuven, Leuven, Belgium
[2]Department of Social Affairs and Health, Ecorys, Rotterdam, Netherlands
[3]Department of Endocrinology, University Hospitals Leuven, Leuven, Belgium
[4]Department of Chronic Diseases, Metabolism and Ageing, University of Leuven, Leuven, Belgium
[5]Department of Clinical and Experimental Endocrinology, University of Leuven, Leuven, Belgium
[6]Faculty of Movement and Rehabilitation Sciences, University of Leuven, Leuven, Belgium

Corresponding Author:
Jan Seghers, PhD
Department of Movement Sciences
University of Leuven
Tervuursevest 101
Leuven,
Belgium
Phone: 32 16329048
Email: jan.seghers@kuleuven.be

Related Article:

This is a corrected version. See correction statement: http://mhealth.jmir.org/2018/3/e10159/

Abstract

Background: Conventional face-to-face weight loss and weight control programs are very labor intensive for both the patient and the provider. It is unclear to what extent conventional programs can be (partially) completed by mobile health (mHealth) apps.

Objective: The aim of this study was to compare the effectiveness of different weight loss programs using a combination of conventional and mobile programs among adults who are overweight (body mass index [BMI]>29 kg/m²).

Methods: A single-blinded randomized controlled trial among obese adults was performed from September 2015 to March 2016. The study took place in Leuven, Belgium. Of the 102 eligible (BMI >29 kg/m²) adults, 81 (79%) completed the study. The three intervention groups consisted of a conventional face-to-face weight loss program, a weight loss app program (app group), and a partial face-to-face and partial app program (combi group). All intervention groups received the same advice from a dietician and a physical activity coach during a 12-week period. The control group did not receive any information during the same period. Primary outcomes were weight reduction (5% decrease of baseline weight in kg), BMI, metabolic risk factors, dietary pattern, and physical activity.

Results: Significant more participants in all three intervention groups lost at least 5% or more of their weight at baseline compared with the control group. No significant difference was found between the combi group and the conventional group. A trend was found that more participants in the combi group lost 5% or more compared with the app group (19%), *P*=.06. A significant time x group effect was found for BMI and metabolic risk factors, with the control group having the worst results and the combi group being significantly better with regard to BMI compared with the app group. No significant group x time effects were found for the intake of different food and drinks and moderate to vigorous physical activity (MVPA).

Conclusions: The results of this study show that a conventional weight loss program could partially be completed with an mHealth program without affecting the effectiveness.

Trial Registration: Clinicaltrials.gov NCT02595671; https://clinicaltrials.gov/ct2/show/NCT02595671 (Archived by WebCite at http://www.webcitation.org/6w1H0x1Q6)

(JMIR Mhealth Uhealth 2018;6(1):e14) doi:10.2196/mhealth.7713

KEYWORDS

obesity; weight loss; mobile applications; diet

Introduction

Background

Obesity remains a serious global health challenge. Approximately 37% (2.1 billion) of the adult world population is overweight or obese, with a prevalence of over 60% in Australia and the United States and between 15% and 30% in Europe [1]. Overweight and obesity were estimated to cause 3.4 million deaths, 3.9% of years of life lost, and 3.8% of disability adjusted life years, globally [1]. Given the still increasing magnitude of the epidemic and the association between excess weight and cardio-metabolic risk factors and comorbidities including diabetes, certain cancers, heart disease, stroke, thrombotic disease, osteoarthritis, sleep apnea, and liver and pulmonary disease [2], apart from prevention, new treatment interventions that reach a wide population are required to address this major public health problem.

Conventional face-to-face weight loss and weight control programs, including components for healthy eating and physical activity, have been found to be effective [3,4]. Unfortunately, these programs were found to be very labor intensive for both the patient and the health professional because of the frequent, lengthy clinic visits (hour-long visits, often weekly, for several months or longer). This places a high burden on the health care professionals and patients. Furthermore, such programs are only effective when patients are committed to invest over a long time period [3].

New Developements

New generations of mobile health (mHealth) technologies that make use of mobile phones or tablets for delivering health information and real-time tailored feedback are emerging and offer good potential for delivery of weight loss programs that are less labor intensive [5]. More specific, there is growing interest into the use of mobile apps to deliver weight loss programs because of their low cost advantages and ability to reach a large number of people because of the increasing number of mobile phone ownership [6-8]. So far, multiple freely available mobile weight loss apps are available at the commercial market (eg, Google Play Store and iTunes App Store). A recent review concluded that computer-tailored and mobile interventions positively affect lifestyle behavior up to 1 year [8]. However, when compared with conventional face-to-face interventions, mHealth interventions seem to be less effective [7]. Moreover, the information quality and evidence-based content of mHealth apps needed improvement [4].

So far, most studies evaluated the effectiveness of a conventional face-to-face weight loss program, an mHealth weight loss program, or a conventional program plus mHealth. It remains unclear to what extent a conventional face-to-face weight loss program could (partially) be completed with a weight loss app. Therefore, the aim of this study was to compare the effectiveness of three weight loss programs (eg, conventional face-to-face weight loss program, a mobile weight loss app [app group], and a partial face-to-face or partial app program [combi group]) with no intervention program among adults with obesity (BMI>29 kg/m²).

Methods

Participants

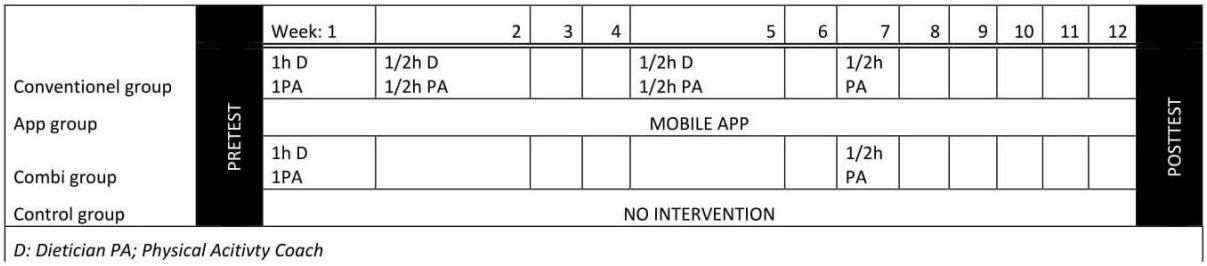
From September 2015 to November 2015, overweight adults living in the Leuven (Belgium) region were recruited for this single-blinded randomized controlled trial (RCT). Inclusion criteria included a body mass index (BMI; calculated as weight in kilograms divided by height in meters squared) between 29 and 34 kg/m² (based on patient metabolic characteristics visiting registered dieticians and qualified physical activity coach in a primary care setting), in the age range of 18 to 65 years, having an email address, and having a personal computer or tablet, or mobile phone. The exclusion criteria were suffering from a known physical (eg, orthopedic limitations and stroke) and/or psychological (eg, eating disorders and depression) disease or comorbidity, intake of any medication with possible impact on body weight, endurance capacity, currently treated for diabetes (both type 1 and 2), sleep apnea determined during the last year, a history of systematic strength or endurance training (moderate to high intensity training more than once a week) in the year before the beginning of the trial, a history of following a supervised dietary advice in the year before the beginning of the trial, having a history of bariatric surgery or any other malabsorption-related disease, and pregnancy.

Recruitment and Randomization

Through flyers, social media, and advertisements in local media, overweight and obese adults were invited to participate in a 12-week weight loss intervention. Every person with an interest in the study was invited to attend a general information session about health risks related to overweight, importance of regular physical activity, healthy eating for successful weight loss, and information about this study. After this session, the invitees could sign up for participation in the study. After signing informed consent, the principle investigator allocated the participants in the different groups by means of random number allocation in Excel (Microsoft; see Figure 1). The allocation rate was 1/1/1/1. Participation in this study was free for all participants. All measurements were taken by a blinded assessor. The study took place, including introduction, pre- and postmeasurements, as well as counseling session, at the premises of the Department of Movement Sciences, Physical Activity, Sports and Health Research Group, KU Leuven in Leuven, Belgium. The study was performed from September 2015 to March 2016. All measurements were conducted by registered dieticians and a qualified physical activity coach. Study procedures were approved by the Medical Ethics Committee

University Hospital Leuven (registration number S57538). This trial was registered at clinicaltrials.gov, number NCT02595671.

Figure 1. Description of intervention.



Interventions

Conventional Face-to-Face Weight Loss Program (Conventional Group)

Participants of this group received an individualized diet plan from a registered dietician. Furthermore, each participant received a personalized physical activity plan for 12 weeks from a physical activity coach. In both plans, behavioral change techniques such as self-monitoring, action planning, and relapse prevention were incorporated [9,10]. In the first week, participants had a 1-hour intake with the dietician and a 1-hour intake with the physical activity coach. In the second and fifth week, participants had face-to-face sessions with the dietician (30 min) and physical activity coach (30 min). In the seventh week, participants received an additional session with the physical activity coach. The main advice with regard to nutrition was to reduce their daily energy intake with 500 kcal, a protein intake at 25% of daily energy intake, keep a low glycemic index, and increase intake of fruit and vegetables. The main advice concerning physical activity was to be physically active on at least 5 days a week (preferably all days a week) for at least 30 min at a moderate to high intensity. Participants in this group did not receive access to the mobile weight loss app.

A Mobile Weight Loss App (App Group)

Participants in this group received an account to use the digital mobile app. The app consisted of six parts: digital advice for their dietary pattern and physical activity, how to challenge themselves, self-monitoring (step count), library with (scientific) information on nutrition and physical activity but also recipes, a help button for advice, and a link to a Facebook group. The app was available for Android and iPhone operating system (iOS, Apple Inc). The content of the digital advice matched with the conventional advice. See Figure 2 for screenshots of the app.

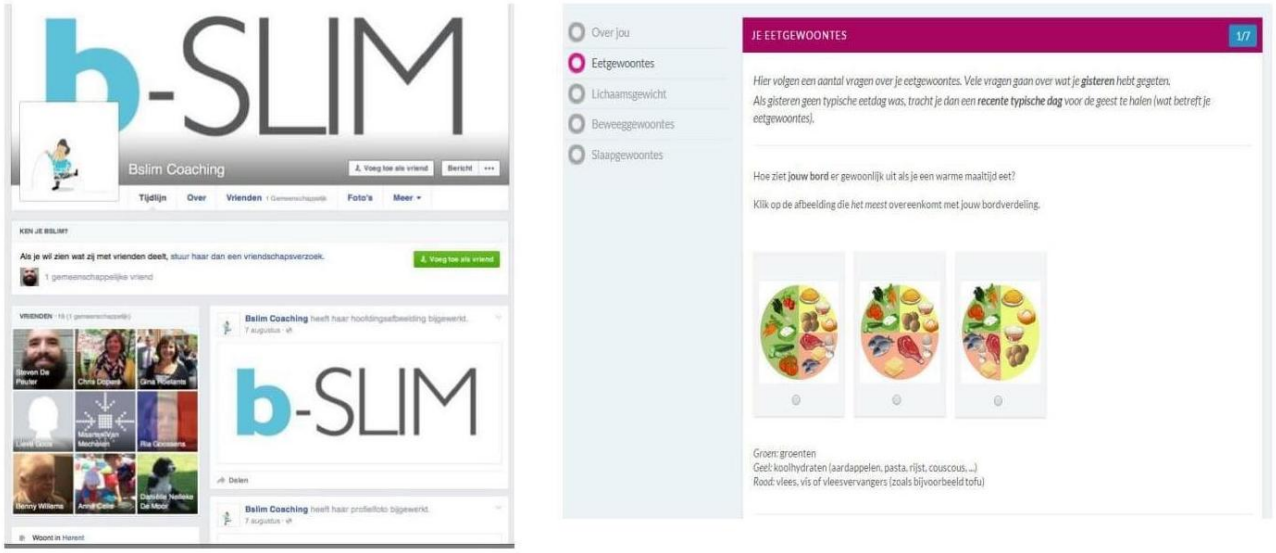
Partial Conventional or Partial Mobile Weight Loss Program (Combi Group)

The subjects of this group received first a 1-hour intake with the dietician and a 1-hour intake with the physical activity coach in the first week. Within these consultations, they received the same information as the conventional treatment group. Additionally, these subjects received an account to use the mobile weight loss app. In the seventh week, participants received an additional face-to-face session with the physical activity coach. Compared with the conventional group, participants received two lesser 30 min counseling sessions with a dietitian and physical activity coach during the intervention.

Control Group

Participants were informed that they were on the waiting list for the weight loss program.

Figure 2. Screenshots of the mobile weight loss app.



Outcome Measurements

Weight Loss

Percentage of participants with at least 5% decrease in baseline weight in kg (5% criterion) [11], BMI (kg/m²) calculated from measured body weight (kg), and body height (m). Weight and height were measured according to the standardized method as described in the World Health Organization, Technical Report Series 854 [12].

Cardio-Metabolic Risk Factors

Waist circumference (WC) was measured with an inelastic tape, placed directly on the skin, perpendicularly to the long axis of the body while the subject stood balanced on both feet, with both arms hanging freely [12]. WC of more than 102 cm for men and more than 88 cm for female was taken as abnormal. Blood pressure (BP) was measured with mercury sphygmomanometer in the sitting position after 5 min of rest. Hypertension was defined as BP>130mmHg for systolic or >85mmHg diastolic or on the basis of hypertension treatment.

High density lipoprotein cholesterol (HDL-C) and triglycerides (TG) were determined using a CardioChek Point-of-Care Self-Test device (Cardiochek PA, Polymer Technology Systems Inc., Indianapolis, IN, United States) [13]. TG 150 mg/dl or more and serum HDL-C less than 40 mg/dl for males and less 50 mg/dl for females was considered abnormal. Glucose levels (average of 2 measurements) were measured by the participants themselves in a fasted state by means of the BGStar measurement (Sanofi). Glucose levels of 100 mg/dl or more was considered abnormal.

Dietary Pattern

A validated digital Food Frequency Questionnaire (FFQ), developed to estimate the overall dietary pattern, was used to measure dietary changes during the 12-week period [14]. A higher change score on the FFQ means a healthier dietary pattern, including more fruit, vegetable, more water, more fish, less soda and alcoholic drinks, and less meat.

Physical Activity Behavior

Physical activity was measured objectively and by means of a self-administrated questionnaire. Objective measurement of physical activity was provided with a tri-axial accelerometer (ActiGraph, model wGT3X-BT, LLC, Pensacola, Florida, United States) [15]. Absolute time spent engaged in moderate (3-5.9 metabolic equivalent of tasks [METS]) and vigorous (≥6METS) intensity activity was calculated (moderate to vigorous physical activity [MVPA]).

The International Physical Activity Questionnaire-Short Form (IPAQ-SF) was used to estimate the amount of self-reported physical activity in the past week [15]. On the basis of these data, participants were categorized into (1) inactive, (2) minimally active, or (3) health enhancing physical activity (HEPA) active [16].

Statistical Analyses

All results were expressed as mean (standard deviation, SD) and mean difference (SD). Intention-to-treat (ITT) analyses were performed. Differences between groups in the baseline data regarding anthropometric (including BMI), dietary patterns, physical activity, and cardio-metabolic risk factors were analyzed using an analysis of variance (ANOVA) comparison test or chi-square test. BMI, dietary pattern, physical activity, and cardio-metabolic risk factors data were analyzed using a 4 × 2 mixed-model repeated-measures ANOVA with group and time (pre vs post) as factors and gender as covariate. Significant interactions were further analyzed by means of Tukey test *post-hoc* analysis. Statistical significance was conventionally considered as $P \leq .05$. All analyses were performed with Statistical Package for the Social Sciences (SPSS) version 23 (IBM Corp).

The sample size for equivalence studies was calculated based on the 5% criterion for weight loss. On the basis of the results of a recent RCT [17], we estimated the success rates of all intervention to be 37%. Together with a power of 0.80 and a significance level of 5%, this led to a needed sample size of 30 per group.

Results

Participants

Of the 122 initially recruited participants, 102 completed the trial (79%; 102/122; see Figure 3). For baseline characteristics see Table 1. There were no baseline differences between the four groups, except for gender (*P*=.02). In the combi group, significant more men were included.

Weight Loss

Significant more participants in all three intervention groups lost at least 5% or more of their weight at baseline compared with the control group (see Figure 4). A trend was found that more participants in the combi group lost 5% or more compared with the app group (19%, *P*=.06). No significant difference was found between the combi group and the conventional group.

A significant time x group effect was found for BMI (*P*=.006), with the control group being significantly different compared with all other intervention groups. No significant decrease was found in the control group. In the conventional group, app group, and combi group, BMI decreased significantly (*P*=.004, *P*=.005, and *P*<.001, respectively; see Table 2).

No significant differences were found between the conventional group and the app group and between the conventional group and the combi group (*P*=.41). However, the combi group had significantly higher decrease in BMI compared with the app group (*P*=.03).

Cardio-Metabolic Risk Factors

A significant time x group effect was found for cardio-metabolic risk factors (*P*=.05,) with the control group being significant worse compared with all other intervention groups. A nonsignificant increase was found in the control group (*P*=.18). Within the conventional group, app group, and combi group, a decrease in metabolic risk factors was found, but this change was not significant (*P*=.12, *P*=.15, and *P*=.23, respectively; see Table 2).

No significant differences were found between the three intervention groups. However, all intervention groups had significant higher decreases in cardio-metabolic risk factors compared with the control group (all *P*<.05).

Dietary Pattern

No significant group x time effect was found for dietary pattern (see Table 2). However, a borderline group x time effect was observed for total energy intake (*P*=.05). All groups reduced their total energy intake; however, only significant changes were found within the conventional group (*P*=.001), app group (*P*=.001), and combi group (*P*<.001) and not in the control group (*P*=.22).

Physical Activity

No significant group x time effects were found for MVPA. Furthermore, no significant changes were found in any of the groups with regard to the percentage of participants that fulfilled the IPAQ minimally active criteria and the HEPA active criteria (see Table 2).

Figure 3. Flowchart of trial.

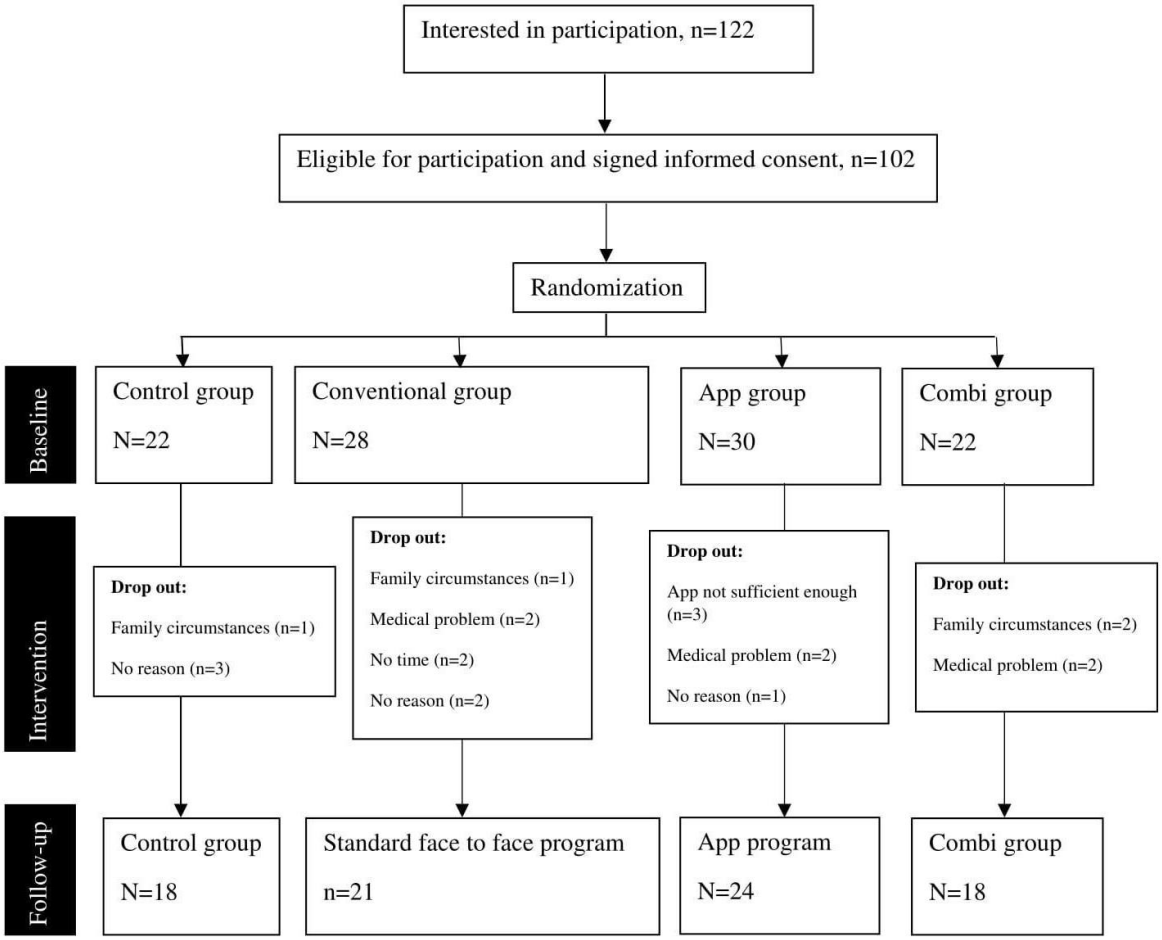


Table 1. Baseline characteristics of the participating adults.

Characteristics	Control group (n=18)	Conventional group (n=21)	App group (n=24)	Combi group (n=18)
Age (years), mean (SD[a])	45 (10.2)	46 (9.2)	44 (12.4)	45 (9.6)
Female (%)	75	84	72	48[b]
Weight (kg), mean (SD)	92 (10.2)	90 (9.1)	90 (10.1)	96 (12.0)

[a]SD: standard deviation.
[b]*P*<.05.

Figure 4. Weight loss (percentage of persons losing 5% of baseline weight).

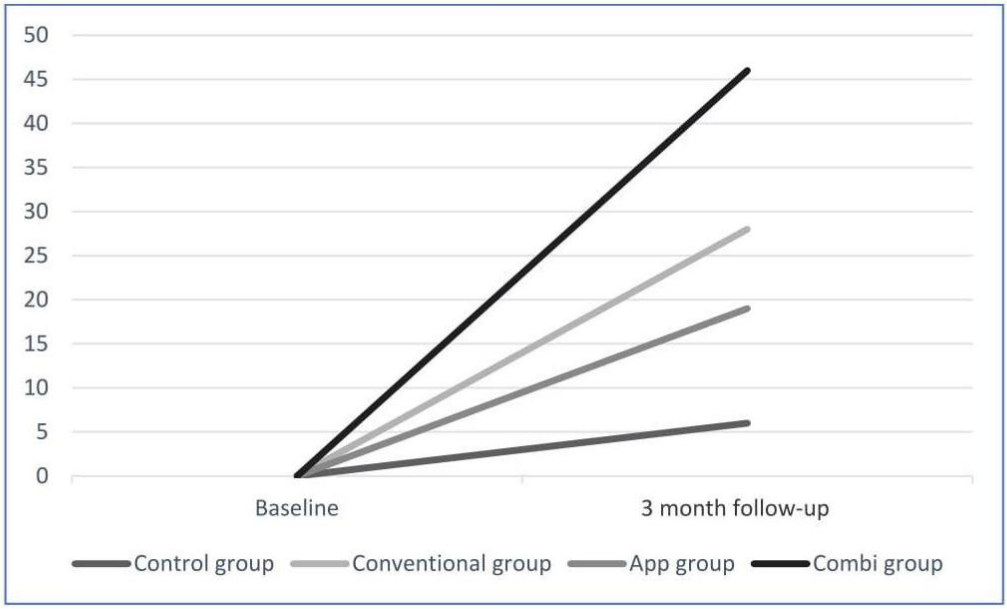


Table 2. Changes in body mass index (BMI), metabolic risk factors, physical activity, and dietary pattern.

Factors	Control group (n=22)		Conventional group (n=28)		App group (n=30)		Combi group (n=22)	
	Pre, mean difference (SD[a]) or n (%)	Post, mean difference (SD)	Pre, mean difference (SD) or n (%)	Post, mean difference (SD)	Pre, mean difference (SD) or n (%)	Post, mean difference (SD)	Pre, mean difference (SD) or n (%)	Post, mean difference (SD)
BMI[b]	32 (2.0)	0.1 (1.0)	32 (2.0)	−1.0 (1.3)	32 (2.1)	−0.7 (1.0)	32 (2.2)	−1.3 (1.2)
Metabolic risk	2.9 (1.2)	0.5 (1.4)	3.0 (1.0)	−0.6 (1.4)	3.2 (1.3)	−0.5 (1.5)	2.9 (1.0)	−0.3 (1.1)
Physical activity								
Category 2	17 (59%)	0 (0.7)	16 (64%)	0.11 (0.6)	14 (58%)	0.05 (0.6)	20 (71%)	0.04 (0.6)
HEPA[c]	6 (21%)	−0.06 (0.4)	7 (28%)	0.0 (0.6)	3 (11%)	−0.04 (0.5)	7 (25%)	−0.08 (0.5)
MVPA[d] (min)	324 (89)	−1.7 (63.7)	314 (82)	11.8 (61.4)	333 (82)	3.6 (72.8)	348 (90)	−33.5 (39.8)
Overall score nutrition pattern	69.5 (13.1)	3.0 (6.8)	69.7 (11.5)	11.,2 (13.7)	71.5 (12.6)	8.1 (13.8)	70.0 (14.9)	8.7 (12.6)
Energy intake (kcal)	1534.2 (548.3)	−115.1 (381.8)	1453.3 (413.6)	−392.7 (302.9)	1489.5 (414.9)	−192.2 (247.4)	1456.5 (397.6)	−287.3 (277.3)

[a]SD: standard deviation.

[b]BMI: body mass index.

[c]HEPA: health enhancing physical activity.

[d]MVPA: moderate to vigorous physical activity.

Discussion

Principal Findings

This study evaluated whether conventional weight loss programs could be (partially) completed with an mHealth app. The results of our study show that when replacing a part of the conventional program by an mHealth app does not affect the effectiveness of the program. Although an mHealth app as a single intervention also showed positive results on BMI and weight reduction, this change was smaller compared with the conventional and combi group.

Comparison With Prior Work

Our results with regard to BMI and weight reduction are in line with previous studies [18-20]. Most studies show that an mHealth app leads to weight reduction. However, the extent of weight loss reduction varies per study. In our study, weight loss was relatively high within the app group. This could be explained by the inclusion of behavioral components such as self-monitoring, avatar possibility (adjustable image of

themselves withouta number of kilograms), action planning, and relapse prevention in the current app, which enables the participants to change their health behavior and maintain this behavioral change during the whole 12-week intervention period. This approach is in contrast with many apps that had not (or not sufficiently) incorporated a behavioral component [5,6].

In our study, metabolic risk factors decreased in our intervention groups. These results are in line with other studies, showing that a combination of a calorie intake reduction combined with physical activity reduces metabolic risk factors [21,22]. However, within the control group, metabolic risk factors increased while they also decreased their energy intake. This could indicate that there is a threshold for the level of energy intake reduction to have an effect on metabolic risk factors.

All groups in our study reduced their energy intake during the trial period. The conventional group and combi group showed highest decrease. Interestingly, the reduced energy intake was accompanied with an overall improvement of their dietary quality. This could be the added value of a health professional. They monitor the patient and provide individualized advice and personalized solutions to certain person specific problems. In an mHealth app, such a personalized approach is not possible to such an extent. Furthermore, this interaction with a health professional might be of high importance for long-term maintenance of the results. Future long-term studies should further evaluate the most effective combination of a health professional and an mHealth app.

Most previous studies showed that conventional weight loss programs, as well as mobile weight loss programs, have a positive effect on the level of physical activity [23]. In our study, no significant improvement was found. This could be explained by the fact that a high percentage of our sample population was already physically active at baseline. More than 60% of our participants reported that they were already active on at least 5 days a week, 30 min on a moderate intensity level or 3 times a week on a high intensity level. The main advice concerning physical activity in our study was to be physically active on at least 5 days a week (preferably all days a week) for at least 30 min at a moderate to high intensity. If participants felt that they were already active enough, it is possible that the goal setting approach to increase their daily level of physical activity could have been marginalized or ignored. Although we have to interpret these results with caution as they are based on a self-reported questionnaire, it could explain why our participants did not significantly increase their physical activity levels. Nonetheless, adults' activity levels have been shown to vary depending on season of measurement, with lower levels of activity in the winter versus spring or summer months [24]. This study was conducted from September 2015 to March 2016, with most participants enrolled by midautumn and followed up through winter. Despite this expected decrease in adult's physical activity levels because of seasonal variation, participants in this study were successful in maintaining their activity level during winter months. This was also the case in the control condition. This may be because regular measurement can encourage participants to reflect on their physical activity [25].

The amount of sessions in our study might be different compared with other similar studies. In our conventional face-to-face weight loss program, the number of sessions with a dietician was based on the standard of care and the number of sessions that are financially reimbursed through the Belgian social security system. Unfortunately, there is currently no financial reimbursement for the sessions with a physical activity coach in Belgium. Therefore, the content and number of sessions in the conventional face-to-face weight loss program were based on their hypothesized effectiveness from previous research among Belgian adults [26]. Compared with the conventional face-to-face weight loss program, the number of face-to-face contacts with a dietician and physical activity coach were reduced in the combi group intervention (partial conventional or partial mobile weight loss program) to offer the participants a feasible blended intervention condition. Furthermore, both the conventional group and the combi group were offered an additional session of 30 min with a professional physical coach in week 7. In both the conventional face-to-face weight loss program, as well as in the combi group, this additional session with a professional physical activity coach halfway through the intervention period (week 7) specifically targeted self-efficacy by using self-regulation techniques such as action planning and relapse prevention techniques (ie, barriers and solutions) that can be applied to all behavior changes (physical activity as well as food intake).

Limitations

Our study has a few limitations that should be kept in mind. The first limitation of this study is the sample size. Although 102 participants started the trial, some participants dropped out during the trial (n=21). However, when using the data of our combi group, a sample size of 15 would have been required. Furthermore, to see whether the data of these dropouts affected our results, a per-protocol and an ITT analyses was performed, which showed no differences on the main outcomes. A second limitation is the use of the accelerometer on the wrist. The data with regard to physical activity levels in our study were relatively high. Previous studies have already shown that the data from the accelerometer depends on the location of measurement on the body [27]. Additionally, the IPAQ has been shown to overestimate the amount of physical activity [28]. Unfortunately, a more accurate device or questionnaire is currently not available. A third limitation is the self-reported glucose measurement. They were instructed to take two consecutive measurements in a fasted state. Whether participants followed these instructions remains unclear. A forth limitation is the duration of our intervention of 3 months. The duration and frequency of the intervention was based on the weight loss programs adults in Belgium would currently receive. In addition, the aim of this study was not to evaluate the difference in long-term effects. Our aim was to initially evaluate whether current conventional care could be replaced by a mobile app. A fifth limitation is the way weight change was measured. In our study, the 5% weight loss criterion was used instead of actual weight lost because the first 5% of weight loss offers the greatest health benefits. Although weight loss expressed in kg would be interesting, the most important outcome is whether

people actually achieve a level where positive effects on health occur.

Conclusions

In conclusion, our study showed that a conventional weight loss program could partially be completed with an mHealth app without affecting the effectiveness. Such combined approach could support health professionals and reduce their workload. Further ways of combining conventional weight loss programs with mHealth apps should be further explored. Whether such combined programs are also cost-effective should be further investigated. Furthermore, long-term studies should evaluate whether the effects of a combined program can be maintained over a long time period.

Acknowledgments

This project is partially funded and realized in collaboration with imec, Belgium. BrandNewHealth developed the weight loss app.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 461KB - mhealth_v6i1e14_app1.pdf]

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781. [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Lavie CJ, De SA, Parto P, Jahangir E, Kokkinos P, Ortega FB, et al. Obesity and prevalence of cardiovascular diseases and prognosis-the obesity paradox updated. Prog Cardiovasc Dis 2016;58(5):537-547. [doi: 10.1016/j.pcad.2016.01.008] [Medline: 26826295]
3. Orozco LJ, Buchleitner AM, Gimenez-Perez G, Roqué IF, Richter B, Mauricio D. Exercise or exercise and diet for preventing type 2 diabetes mellitus. Cochrane Database Syst Rev 2008 Jul 16(3):CD003054. [doi: 10.1002/14651858.CD003054.pub3] [Medline: 18646086]
4. Leslie WS, Taylor R, Harris L, Lean ME. Weight losses with low-energy formula diets in obese patients with and without type 2 diabetes: systematic review and meta-analysis. Int J Obes (Lond) 2017 Jan;41(1):96-101. [doi: 10.1038/ijo.2016.175] [Medline: 27698345]
5. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
6. Okorodudu DE, Bosworth HB, Corsino L. Innovative interventions to promote behavioral change in overweight or obese individuals: a review of the literature. Ann Med 2015 May;47(3):179-185 [FREE Full text] [doi: 10.3109/07853890.2014.931102] [Medline: 25011006]
7. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012(8):CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
8. Afshin A, Babalola D, Mclean M, Yu Z, Ma W, Chen C, et al. Information technology and lifestyle: a systematic evaluation of internet and mobile interventions for improving diet, physical activity, obesity, tobacco, and alcohol use. J Am Heart Assoc 2016 Aug 31;5(9) [FREE Full text] [doi: 10.1161/JAHA.115.003058] [Medline: 27581172]
9. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
10. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
11. Wing RR, Lang W, Wadden TA, Safford M, Knowler WC, Bertoni AG, et al. Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes. Diabetes Care 2011 Jul;34(7):1481-1486 [FREE Full text] [doi: 10.2337/dc10-2415] [Medline: 21593294]
12. WHO. Physical status: the use and interpretation of anthropometry URL: http://www.who.int/childgrowth/publications/physical_status/en/ [accessed 2017-12-21] [WebCite Cache ID 6vs8bW4YV]
13. Williams AD, Ahuja KD, Brickwood KJ, Fell J. Evaluation of the cardiochek portable whole blood analyses for use in the fitness industry. J Exerc Physiol 2011;14:71.

14. Matthys C, Meulemans A, Van der Schueren B. Development and validation of general FFQ for use in clinical practice. Ann Nutr Metab 2015;67:239.
15. John D, Freedson P. ActiGraph and Actical physical activity monitors: a peek under the hood. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S86-S89 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399f5e] [Medline: 22157779]
16. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
17. Spring B, Duncan JM, Janke EA, Kozak AT, McFadden HG, DeMott A, et al. Integrating technology into standard weight loss treatment: a randomized controlled trial. JAMA Intern Med 2013 Jan 28;173(2):105-111 [FREE Full text] [doi: 10.1001/jamainternmed.2013.1221] [Medline: 23229890]
18. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
19. Felix HC, West DS. Effectiveness of weight loss interventions for obese older adults. Am J Health Promot 2013;27(3):191-199 [FREE Full text] [doi: 10.4278/ajhp.110617-LIT-259] [Medline: 23286596]
20. Anderson LM, Quinn TA, Glanz K, Ramirez G, Kahwati LC, Johnson DB, Task Force on Community Preventive Services. The effectiveness of worksite nutrition and physical activity interventions for controlling employee overweight and obesity: a systematic review. Am J Prev Med 2009 Oct;37(4):340-357. [doi: 10.1016/j.amepre.2009.07.003] [Medline: 19765507]
21. Balk EM, Earley A, Raman G, Avendano EA, Pittas AG, Remington PL. Combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: a systematic review for the community preventive services task force. Ann Intern Med 2015 Sep 15;163(6):437-451 [FREE Full text] [doi: 10.7326/M15-0452] [Medline: 26167912]
22. Normandin E, Chmelo E, Lyles MF, Marsh AP, Nicklas BJ. Effect of resistance training and caloric restriction on the metabolic syndrome. Med Sci Sports Exerc 2017 Mar;49(3):413-419. [doi: 10.1249/MSS.0000000000001122] [Medline: 27741216]
23. Laddu D, Dow C, Hingle M, Thomson C, Going S. A review of evidence-based strategies to treat obesity in adults. Nutr Clin Pract 2011 Oct;26(5):512-525. [doi: 10.1177/0884533611418335] [Medline: 21947634]
24. Shephard RJ, Aoyagi Y. Seasonal variations in physical activity and implications for human health. Eur J Appl Physiol 2009 Oct;107(3):251-271. [doi: 10.1007/s00421-009-1127-1] [Medline: 19609553]
25. Waters L, Reeves M, Fjeldsoe B, Eakin E. Control group improvements in physical activity intervention trials and possible explanatory factors: a systematic review. J Phys Act Health 2012 Aug;9(6):884-895. [Medline: 22898467]
26. Van HA, Delecluse C, Bogaerts A, Boen F. The long-term effectiveness of need-supportive physical activity counseling compared with a standard referral in sedentary older adults. J Aging Phys Act 2014 Apr;22(2):186-198. [doi: 10.1123/japa.2012-0261] [Medline: 23628840]
27. Shiroma EJ, Schepps MA, Harezlak J, Chen KY, Matthews CE, Koster A, et al. Daily physical activity patterns from hip- and wrist-worn accelerometers. Physiol Meas 2016 Sep 21;37(10):1852-1861. [doi: 10.1088/0967-3334/37/10/1852] [Medline: 27654140]
28. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]

Abbreviations

ANOVA: analysis of variance
BMI: body mass index
BP: blood pressure
FFQ: Food Frequency Questionnaire
HDL-C: high density lipoprotein cholesterol
HEPA: health enhancing physical activity
IPAQ-SF: International Physical Activity Questionnaire-Short Form
ITT: intention-to-treat
METS: metabolic equivalent of tasks
mHealth: mobile health
MVPA: moderate to vigorous physical activity
RCT: randomized controlled trial
SD: standard deviation
TG: triglycerides
WC: waist circumference

Edited by G Eysenbach; submitted 18.03.17; peer-reviewed by G Flores Mateo, B Nezami, F Saigí-Rubió; comments to author 10.08.17; revised version received 03.10.17; accepted 16.11.17; published 11.01.18.

Please cite as:
Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J
Face-to-Face Versus Mobile Versus Blended Weight Loss Program: Randomized Clinical Trial
JMIR Mhealth Uhealth 2018;6(1):e14
URL: http://mhealth.jmir.org/2018/1/e14/
doi:10.2196/mhealth.7713
PMID:29326093

Original Paper

A Wearable Sensor-Based Exercise Biofeedback System: Mixed Methods Evaluation of Formulift

Martin Aidan O'Reilly[1,2], PhD, BEng (Hons); Patrick Slevin[1,2], MSc; Tomas Ward[3,4], PhD; Brian Caulfield[1,2], PhD

[1]Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland

[2]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland

[3]Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[4]School of Computing, Dublin City University, Dublin, Ireland

Corresponding Author:
Martin Aidan O'Reilly, PhD, BEng (Hons)
Insight Centre for Data Analytics
University College Dublin
O'Brien Centre for Science, 3rd Fl
Belfield
Dublin, D4
Ireland
Phone: 353 871245972
Email: martin.oreilly@insight-centre.org

Abstract

Background: *Formulift* is a newly developed mobile health (mHealth) app that connects to a single inertial measurement unit (IMU) worn on the left thigh. The IMU captures users' movements as they exercise, and the app analyzes the data to count repetitions in real time and classify users' exercise technique. The app also offers feedback and guidance to users on exercising safely and effectively.

Objective: The aim of this study was to assess the *Formulift* system with three different and realistic types of potential users (beginner gym-goers, experienced gym-goers, and qualified strength and conditioning [S&C] coaches) under a number of categories: (1) usability, (2) functionality, (3) the perceived impact of the system, and (4) the subjective quality of the system. It was also desired to discover suggestions for future improvements to the system.

Methods: A total of 15 healthy volunteers participated (12 males; 3 females; age: 23.8 years [SD 1.80]; height: 1.79 m [SD 0.07], body mass: 78.4 kg [SD 9.6]). Five participants were beginner gym-goers, 5 were experienced gym-goers, and 5 were qualified and practicing S&C coaches. IMU data were first collected from each participant to create individualized exercise classifiers for them. They then completed a number of nonexercise-related tasks with the app. Following this, a workout was completed using the system, involving squats, deadlifts, lunges, and single-leg squats. Participants were then interviewed about their user experience and completed the System Usability Scale (SUS) and the user version of the Mobile Application Rating Scale (uMARS). Thematic analysis was completed on all interview transcripts, and survey results were analyzed.

Results: Qualitative and quantitative analysis found the system has "good" to "excellent" usability. The system achieved a mean (SD) SUS usability score of 79.2 (8.8). Functionality was also deemed to be good, with many users reporting positively on the systems repetition counting, technique classification, and feedback. A number of bugs were found, and other suggested changes to the system were also made. The overall subjective quality of the app was good, with a median star rating of 4 out of 5 (interquartile range, IQR: 3-5). Participants also reported that the system would aid their technique, provide motivation, reassure them, and help them avoid injury.

Conclusions: This study demonstrated an overall positive evaluation of *Formulift* in the categories of usability, functionality, perceived impact, and subjective quality. Users also suggested a number of changes for future iterations of the system. These findings are the first of their kind and show great promise for wearable sensor-based exercise biofeedback systems.

(JMIR Mhealth Uhealth 2018;6(1):e33) doi:10.2196/mhealth.8115

KEYWORDS

mHealth; feedback; posture; exercise therapy; biomedical technology; lower extremity; physical therapy specialty

Introduction

Background

Resistance training is an exercise modality used in rehabilitation and strength and conditioning (S&C) settings. Adhering to a resistance training exercise program can increase a person's muscular strength, hypertrophy, and power [1]. This can improve their sporting performance, mood, and quality of life [2,3]. However, many people completing exercise programs encounter various difficulties when performing their exercises without the supervision of a trained exercise professional such as an S&C coach. One such difficulty is that in such circumstances, people may execute their exercises incorrectly [4,5]. Incorrect alignment during exercise, incorrect speed of movement, and poor quality of movement may have an impact on the efficacy of exercise and may therefore result in a poor outcome [4,5]. Exercising with aberrant biomechanics may also heighten one's risk of injury [6], necessitating technological solutions to provide accurate assessment of exercise form.

Inertial measurement units (IMUs) have been shown recently to be an accurate method for such exercise assessment. Wearable IMUs are able to acquire data pertaining to the linear and angular motion of limb segments and can also be used to measure a body's three-dimensional orientation [7,8]. They are small, inexpensive, easy to set up, and facilitate the acquisition of human movement data in unconstrained environments [9]. Recent research has shown that a diverse range of exercises can be accurately evaluated with multiple and individual IMU setups [10-13]. These range from early stage rehabilitation exercises such as heel slides and straight leg raises [12] to more complex S&C exercises such as bodyweight squats [14], lunges [15], and single-leg squats [13,16,17]. More cost-effective and practical systems using a single body-worn IMU have also been shown to be effective in the analysis of exercise technique [13,15,18,19]. Such single IMU-based systems are considered preferential where they can provide equivalent exercise analysis quality to multiple IMU setups. Recently, it has been shown that for the detection of acute, naturally occurring deviations in compound lower-limb S&C exercises such as the barbell squat and the deadlift, personalized classification systems are superior in accuracy to the global ones [20]. It has also been shown that such personalized systems enable a single IMU to accurately classify repetitions of such exercises as "acceptable" or "aberrant" [20].

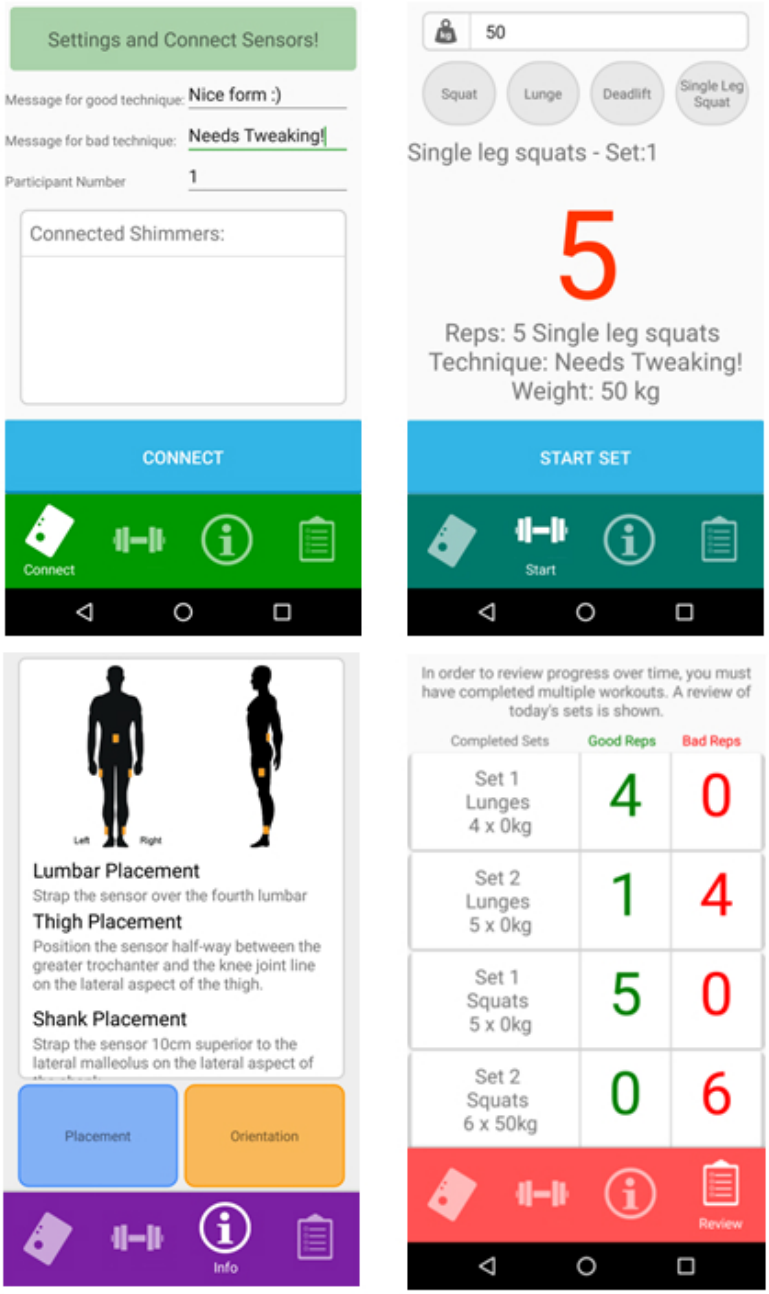
Although all the aforementioned work demonstrates the technological proficiency of IMU-based exercise biofeedback systems in classifying exercise technique, little is currently known about the user experience and users' perceptions of such biofeedback systems. There is currently a surge of usability and system evaluation studies being published in the mobile health (mHealth) and ubiquitous health (uHealth) field [21-24]; however, there is a sparsity of such studies pertaining to IMU-based exercise biofeedback systems. Some past work has assessed the usability of IMU-based exercise biofeedback systems [25,26] but, to the author's knowledge, there has not yet been any evaluation studies of biofeedback systems that classify exercise quality based on data from a wearable sensor and relay feedback to users via a mobile app.

The study aimed to evaluate a recently developed IMU-based exercise biofeedback system called *Formulift*. *Formulift* consists of a mobile app and a single Shimmer IMU (Shimmer, Dublin, Ireland). The IMU is worn on the user's left thigh and tracks their motion as they complete the following four exercises: squats, single-leg squats, lunges, and deadlifts. The mobile app processes the signals from the IMU, counts repetitions, and utilizes personalized classification methods to determine if each repetition completed of an exercise is "acceptable" or "aberrant." During a set, the exerciser receives real-time feedback on the completion of repetitions of an exercise; this includes a vibration of the phone and an on-screen repetition counter. The user then receives feedback on their exercise technique following the completion of each set of an exercise. This feedback is shown in Figure 1 (top and bottom right) whereby after a set the exerciser is given a color-coded number indicating their technique quality and a message to reinforce this. The exerciser may then view how many repetitions were completed with acceptable and aberrant form on the review screen. The app also displays a pop-up message if two sets of the same exercise are completed sequentially with aberrant form. This message suggests they seek support of an exercise professional to identify and address their specific movement inefficiency. The app contains instructional information on how to do the exercises with acceptable technique and the option to review a workout session. However, these videos are not specific to the identified aberrant movements. A video comprehensively detailing the system can be seen in Multimedia Appendix 1. A number of screenshots from the app are shown in Figure 1.

Objectives

The aim was to assess the system under a number of categories: (1) usability: the extent to which the system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use; (2) functionality: the ability of an interface or device to perform according to a specifically defined set of parameters [27] whereby the key functions of *Formulift* are to accurately detect and count repetitions of the exercises under study, determine if each repetition was completed with "acceptable" or "aberrant" technique, and provide the user with interpretable feedback on their completed exercise; (3) the perceived impact of the system; and (4) the subjective quality of the system. It was also desired to uncover suggested future improvements to the system. Three different realistic types of system users were employed to complete this evaluation: beginner gym-goers (<6 months experience), experienced gym-goers (>2 years' experience), and qualified S&C coaches. Employing these three types of end users was hypothesized to enable a more comprehensive user-centered design approach to creating future iterations of the *Formulift* system.

Figure 1. The *Formulift* app. User preferences and connecting to the inertial measurement unit (IMU; top left), real-time exercise biofeedback (top right), information and instructions (bottom left), and workout review (bottom right).



Methods

Participants

A total of 15 healthy volunteers participated (12 males, 3 females, age: 23.8 years [standard deviation, SD 1.8], height: 1.79 m [SD 0.07], body mass: 78.4 kg [SD 9.6]). Group 1 included 5 beginner gym-goers with fewer than 6 months experience with resistance training and the exercises used in this study. Group 2 included 5 experienced gym-goers with a minimum of 2 years' experience with resistance training and the exercises used in this study. The final group of system evaluators included practicing S&C coaches with qualifications from the National Strength and Conditioning Association or the United Kingdom Strength and Conditioning Association. Sample size numbers were chosen based on a combination of standard practice for completing qualitative usability studies [28] and in keeping with recent publications that also utilized the quantitative surveys used in this work [29,30]. No participant had a current or recent musculoskeletal injury that would impair their exercise performance. Each participant signed a consent form before completing the study. The University College Dublin Human Research Ethics Committee approved the study protocol.

Data Collection

System Use

The testing protocol was explained to participants upon their arrival to the research laboratory. All participants completed a 5-min warm-up on an exercise bike; during which they were required to maintain a power output of 100 W and cadence of

75 to 85 revolutions per min. Following the warm-up, an investigator positioned a single IMU (Shimmer, Shimmer research, Dublin, Ireland) on the participant at the midpoint of the left femur (determined as halfway between the greater trochanter and lateral femoral condyle).

Video and IMU data were then simultaneously collected as the users completed the four following exercises: bodyweight left leg single-leg squats, bodyweight lunges, bodyweight or barbell squats, and barbell deadlifts. A total of 40 repetitions of each exercise were collected; 20 repetitions were completed with "acceptable" form, and 20 repetitions were completed with "aberrant" form. The "aberrant" repetitions from the 5 beginners were naturally occurring, whereas the 10 experienced participants deliberately induced their "aberrant' form. Following this data collection, the IMU was removed from the participant's left thigh.

The exercise professional then used the segmented videos to label each repetition of the four exercises as being "acceptable" or "aberrant" technique. Four binary random forests classifiers were then created for each participant, each pertaining to one of the four aforementioned exercises. These random forests objects were imported into the biofeedback app to make a personalized exercise classification system for each participant (Figure 1; Multimedia Appendix 1). While their personalized system was created, the participants completed a set of nonexercise-based tasks within the app. Multimedia Appendix 2 is the sheet given to participants listing the tasks that were to be completed "before the exercise analysis session." They involved app navigation, interpretation of information within the app, and following instructions on system use and how to do the exercises. A full description of the exercise biofeedback system can be seen in the attached video (Multimedia Appendix 1).

Following the creation of their personalized biofeedback system, the participant first secured the IMU to their left thigh and then completed the list of tasks outlined in the "during exercise analysis section" of Multimedia Appendix 2. They first connected the wireless Shimmer IMU to the mobile app. They then completed two sets of ten repetitions for each of the four exercises. In the first set of each exercise, they were instructed to exercise with their best possible technique, and in the second, they were asked to try and replicate the mistake they had made before the exercise professional's coaching. Throughout the session they were able to navigate to any point within the app including the "review tab" and to view any instructional content. The whole exercise session was observed by the investigator, who took note of any system crashes and the associated conditions, as the participant used the system and completed their exercises. The session was also simultaneously videoed for review following data collection. Upon completing the required exercises, participants were provided with the opportunity to test any other tasks within the app they desired. Participants then moved on to evaluating the system whereby they were administered surveys and partook in an interview.

Interviews

Immediately after completion of their exercise session with the system, a semistructured interview was conducted with each participant. A Dell Inspiron 5100 laptop (DELL, Texas, US) was used to video-record the interview. The webcam also captured the screen of the Android smartphone, allowing users to demonstrate any specific aspects of the app they wished to discuss. Each interview followed a topic guide to ensure consistent questioning across every interview [31]. This guide can be seen in Multimedia Appendix 3. Open-ended questions were used to garner participant's views and experiences of the system in relation to usability, functionality, and perceived impact. Furthermore, participant's reflections regarding their general evaluation of the system and suggested future changes were also captured.

Surveys

In addition to the interviews, the system was also assessed quantitatively utilizing two surveys. By mixing both quantitative and qualitative research and data, gains in breadth and depth of understanding and corroboration can be achieved, while offsetting the weaknesses inherent to using each approach separately.

The System Usability Scale (SUS) is a short, 10-point questionnaire that has been widely adopted in many domains as a fast and reliable measure of a system's usability. The scale produces a usability score out of 100 (not a percentage) for every user who completes it. These scores can then be compared with the large body of published data on systems assessed with the SUS to find adjective and percentile rankings of a system's usability [29,32].

The "user version of the Mobile Application Rating Scale" (uMARS) was also completed by all participants [22]. This is an adapted version of the "Mobile Application Rating Scale" and is more appropriate for end users of mobile apps [30]. It assesses the app under the areas of engagement, functionality, aesthetics, and information to produce an overall app quality score out of 5. The app's subjective quality and perceived impact are also assessed separately. The perceived impact section of the survey was tailored to this study to investigate the app's perceived impact on a person "exercising with their best technique." No further adaptations were made to the uMARS survey for this work.

Data Analysis

Qualitative

Interview recordings for all participants were transcribed verbatim and anonymized. A grounded-theory approach was then taken to the thematic analysis of the interview transcripts [33,34]. The interview topic guide was used to create an initial coding frame that was then refined as more data were analyzed. Data analysis was conducted by authors MOR and PS. Analysis involved scrutinizing the data to identify patterns, assigning codes to the data, and building themes and subthemes from the codes [35]. To maximize rigor and ensure the reliability of the theme extraction process, the researchers (MOR and PS) met frequently to evaluate the consistency of emergent themes and subthemes, which were further cross-checked for consistency across the three participant types (beginner gym-goers, experienced gym-goers, and S&C coaches) [36]. Discrepancies that were identified during these meetings were resolved through

discussion between researchers, MOR and PS, until an agreement was reached. Data saturation was determined when no new data and no new themes and relationships among the interview data were emerging [37]. A table of themes and subthemes was created with associated quotes. This can be seen in Multimedia Appendix 4 of this paper.

Quantitative

The SUS score was computed for each participant following standard scoring methodology [38]. The mean and SD for the SUS scores was calculated for all participants and for each subgroup (beginner gym-goers, experienced gym-goers, and S&C coaches). The uMARS was also scored following standard procedures [22]. For each participant, a score out of 5 was calculated for engagement, functionality, aesthetics, and information. The mean of these four scores produced an "overall app quality score" for each app user. App subjective quality was quantitatively assessed taking each user's star rating of the app. A perceived impact score, out of 5, was also found for each participant. The means and SDs of all the above scores were found for all participants and the three aforementioned subgroups.

Results

Summary

The *Formulift* system was assessed across four distinct domains: usability, functionality, perceived impact, and overall quality. In the upcoming subsections, results will be presented from both the quantitative surveys and qualitative interviews. Finally, suggested future changes will also be described.

User Version of the Mobile App Rating Scale

The uMARS provided quantitative results on a number of key aspects of the app. A summary of results from the uMARS are summarized in Table 1. This table is referred to throughout the Results section.

Usability and Functionality

The system achieved a mean SUS usability score of 79.2 (SD 8.8). Beginners deemed the system most usable with a score of 86.25 (SD 1.9), whereas the experienced gym-goers and S&C coaches scored the system at 75.5 (SD 9.1) and 74.5 (SD 8.0), respectively. These usability scores put the system at an 85% to 95% percentile based on all published research using the SUS [29] and would deem the system's usability good to excellent on an adjective rating scale [32]. The functionality section of the uMARS also demonstrated an overall positive usability and functionality experience for users with a mean score of 4.2 (SD 0.37; Table 1). Although these surveys demonstrated that *Formulift* was deemed to have good usability, they provided limited insight in to the reasons for this and to what can be improved. This was found in analysis of the interview transcripts as described below.

Three key areas emerged from the interview data in relation to system usability: overall ease of use, the app's interface, and the IMU. In terms of overall ease of use, 14 out of 15 participants remarked on the system being "easy to use," "straightforward," and "intuitive." Example statements included the following:

> *I thought it was so easy to use...I just like how accessible it is as well.* [Beginner gym-goer]
>
> *Very, very easy to use. Really straightforward. You know, easy to get around and realise what you're doing.* [Experienced gym-goer]
>
> *It's very easy to navigate through. It's pretty easy to be honest...It's monkey see, monkey do really.* [S&C coach]

Many participants commented on the intuitive nature of the app. For instance, the layout of the app was acknowledged to be very easy to follow with large icons, large buttons and a minimal number of menus being cited as the reasons for this. Example statements included the following:

> *I mean, it's quite user friendly. The interface; there's not too much going on, on the screen. It's very clear where the info tab is, where the exercise tab is etc.* [S&C coach]
>
> *Large buttons made that easy. It might come in to play more if you've got sweaty hands, but yeah in terms of navigation it was good. The size of the text and the buttons etc. is good. Overall, very good.* [S&C coach]
>
> *Yeah, the UI is really simple. Some other fitness apps are horrific. I hate using them, because they look horrible.* [Beginner gym-goer]
>
> *It was really easy to find things and navigate through.* [Experienced gym-goer]

None of the users reported any difficulties interpreting the language used within the app. The color used within the app was also referred to in a positive manner. It was considered to "make things stand out great" during a session, make the app "attractive," and the color of the repetition number (red, orange, or green) following a set was said to be very useful:

> *I think the three color, "green, orange and red" feedback was a really useful function as it let you know if you're doing something well, something a little bit off or doing something badly.* [Beginner gym-goer]

Participants also offered positive feedback regarding the "How to wear the sensor" section. They found the instructions were very clear and easy to follow:

> *One thing that I thought was done well was just showing you how to place the sensor as well. That could be a big obstacle, if it wasn't shown properly. It would hinder people's ability to use it. It was done well.* [S&C coach]
>
> *You go in to sensor placement/orientation you can't go wrong there. If you do, you have an issue.* [S&C coach]

Table 1. Results from user version of the Mobile Application Rating Scale (uMARS) survey for beginner gym-goers, experienced gym-goers, and strength and conditioning (S&C) coaches. Overall quality is computed as described in the study by Stoyanov et al.

uMARS Section	Beginners (n=5) Mean (SD)	Experienced (n=5) Mean (SD)	S&C (n=5) Mean (SD)	All (n=15) Mean (SD)
Engagement	3.78 (0.48)	3.5 (0.42)	3.67 (0.33)	3.66 (0.42)
Functionality	4.27 (0.22)	4.24 (0.45)	4.08 (0.39)	4.2 (0.37)
Aesthetics	3.78 (0.50)	3.87 (0.81)	4.2 (0.34)	3.9 (0.62)
Information	4.29 (0.58)	4.2 (0.29)	3.9 (0.56)	4.14 (0.53)
Overall quality	4.03 (0.25)	3.95 (0.29)	3.96 (3.96)	3.98 (0.21)
Perceived impact	4.57 (0.39)	4.12 (0.45)	3.28 (0.52)	4.03 (0.70)
Star rating	3.83 (0.68)	4.0 (0.63)	3.6 (0.49)	3.8 (0.63)

In addition to this, a participant spoke positively about wearing the IMU:

> *I'm not conscience about wearing it, nobody can see it, it doesn't feel weighty or anything like that. I almost forget it's on my leg while I'm talking to you.* [Beginner gym-goer]

Participants reported a number of usability issues. The most reported usability issue related to app navigation, in particular, to going back a step within the app. Four participants, who are usually iPhone users, struggled initially to know how to navigate backwards in the app:

> *Maybe as I'm coming from IOS to Android but there was no clear back button so you have to switch in and out or use the phone's button. On an iPhone, there'd always be something on the screen. That was one thing.* [Beginner gym-goer]

> *Just because I'm not used to using android, I didn't know how to go back a step but other than that, no the app itself is very easy to navigate.* [Experienced gym-goer]

Some participants also highlighted the need for more status indicators as a usability issue. Particularly, they highlighted that within the "How to use the App" instructions, there was no on-screen indicator when one reached the last instruction, which meant they did not know that the final instruction had been reached. More importantly, the need for a loading indicator was highlighted when a user pressed the "Analyse my Set" button. One beginner gym-goer commented:

> *...when you don't do that, I will impatiently tap the same button until something happens which in this case caused crashes.* [Beginner gym-goer]

In fact, this crash, caused by multiple taps of the "Analyse my Set" button was one of the most reported functionality issues. Although many users reported no bugs in the app, 5 users expressed experiencing a crash of the same manner. Two other critical bugs were found within the app that caused system crashes. The first was recorded by 4 users who reported a crash when quickly clicking through the "How to use the App" instructions:

> *There is a way of crashing it (the app). If you use the "how to use app menu" and go quickly through the menu, it's pretty easy to crash. It seems like the second time it happens. You can scan through it the first time but not the second.* [Beginner gym-goer]

> *In the app instructions, I was tapping through quickly and it just crashed. I wasn't mashing the button but I was pressing it reasonably quickly.* [Experienced gym-goer]

The second was experienced by 2 users who also found that the app crashed when they quickly navigated between the four main tabs of the app (connect, exercise, information, and review):

> *When I was exiting the app, when I had been looking at the exercises, it was just coming out and hitting all the buttons (demoes bashing all the menu buttons quickly) and the app crashed.* [S&C coach]

These were both programming bugs, which will be amended in future versions of the app. The most recurrent, nonfatal functionality issue mentioned by users regarded the real-time repetition counting during sets. Eight participants described thinking there was a lag at the start of the set, and after 2 or 3 repetitions, it was as if the app caught up and started counting them properly:

> *When I did the first rep of each set, I wasn't sure if it was recording it, until I did the second rep. It would then say "2." Sometimes it would take a couple seconds just to vibrate and register that I'd completed the repetition.* [Beginner gym-goer]

> *The rep counting was also a little bit slow at the start.* [Experienced gym-goer]

However, all participants felt the total repetition count was always correct. Participants also felt the binary classification of exercise repetitions (as acceptable or aberrant technique) was accurate. The beginners and experienced athletes found the system's feedback useful. Gym-goers remarked the following:

> *It was really interesting how it could pick up on the bad ones and I know there were definitely some bad ones in there!* [Beginner gym-goer]

> *I usually am very aware of my form for sets but there was a set of single leg squats where I didn't do the exercise well enough, and the app told me that I hadn't, and I wasn't aware of that but then when I*

> *thought about it the app was definitely right.* [Experienced gym-goer]

The S&C coaches, whose experience and knowledge allowed them to gain more insight to the accuracy of the system, were predominantly content with the system's accuracy. Two S&C coaches did, however, feel the system misclassified a small number of specific repetitions:

> *It worked for everything except my single leg squats I'd say and maybe a little on the lunges.* [S&C coach]

> *Maybe one thing that it wasn't able to discriminate on that well was the last set I did of shallow bodyweight squats. Maybe the accuracy fell off if I was doing something between a ¼ squat and a proper full squat. That was the only one that was a tiny bit inaccurate.* [S&C coach]

Overall, the SUS results, the uMARS, and the thematic analysis have demonstrated the system was usable and functional. The thematic analysis of the interview transcriptions has also uncovered a number of specific functionality issues and aspects of usability that can be amended or improved in future iterations of the app.

Perceived Impact

The quantitative analysis of the perceived impact of the system, through the uMARS, demonstrated that the system was very beneficial to gym-goers in heightening their awareness of, advancing their knowledge of, increasing their motivation to, and their likelihood to seek help with "exercising with best possible technique" (Table 1). Thematic analysis of the interview transcriptions verified these quantitative findings and also uncovered a number of other perceived benefits and disadvantages to use of the system.

All users reported that using the system would aid their technique while exercising. Beginners often mentioned that the system would enable them to learn proper technique, whereas experienced gym-goers stated that the system would be useful particularly when they lose focus or increase the weight they are lifting, and S&C coaches thought the system would help people correct their technique and avoid injury while exercising. Statements included the following:

> *It's also nice to have the feedback on how I'm actually doing things. Personally, when I go to the gym, I may even do a whole workout and not know if things have gone correctly. It's pretty annoying to go home and be thinking, "Did I do my squats right today?," "I'm not actually sure."* [Beginner gym-goer]

> *For people who are just starting out with workout programs and need technique and form, it's helpful. It's helpful also for advanced weightlifting individuals who are looking to prevent injury and that kind of thing I would say.* [Experienced gym-goer]

> *A lot of the glaring issues people have when staring weight training are addressed. If people even just think about 1 or 2 of the issues that the app lays out then their technique can improve immensely in a very short amount of time just from these little bits of information.* [S&C coach]

> *Well advantages would be, obviously you're avoiding injury as you go to the gym. This gives you a new source that can tell you if you're doing it right or wrong or not.* [S&C coach]

Eight users also suggested that use of the system would have a positive effect on their focus and motivation to exercise with acceptable technique. S&C coaches also suggested that the system would be particularly useful in a team setting where athletes sometimes don't process guidance properly or lose focus. All three test groups made statements regarding the system heightening focus, concentration, and/or motivation, such as:

> *It would be a motivational thing as well as obviously the benefit of getting help to correct yourself when you exercise poorly if needed.* [Beginner gym-goer]

> *Particularly, with me, when I'm sometimes doing weights I lose focus, so it would help me keep track.* [Experienced gym-goer]

> *You have definitely got some players where the information goes in one ear and straight out the other. So it would be good for us in the sense that we could connect this up, we analyze what we want to know and they find out straight away if they've done a good or bad rep.* [S&C coach]

Three out of six of the beginner gym-goers also spoke about the system as a tool to build their confidence in how they are exercising. They spoke of the app as a method to boost their likelihood of seeking help from a friend or trainer, a way to get over the initial anxiety of going to a gym, and to reassure them that they are exercising properly. One beginner spoke extensively of this, including saying:

> *Also, having something on my leg is really reassuring because I've always found that with fitness apps on my phone that direct me to exercise, I almost feel like all the information there can be interpreted wrong and when I go to do the exercises I might be misinterpreting them. But whatever it is, just having this on my leg just makes me feel a little bit more confident in doing them and interpreting the information that is provided by the app.* [Beginner gym-goer]

In addition to giving people confidence in their training, 3 S&C coaches felt one of the key benefits to the system is that it would boost people's likelihood to simply start and commit to an exercise program. One of the S&C coaches in fact saw this as the biggest benefit to the system, and another spoke very positively of this aspect of system use:

> *The biggest benefit is it gets people in to the gym.* [S&C coach]

> *I think downloading the app could give a lot of people confidence to walk in to the gym in the first place, that's really, really good.* [S&C coach]

The aforementioned subthemes of the perceived impact of system use, namely, (1) improving technique, (2) increasing

motivation and focus, and (3) promoting participation in exercise, are all well accepted benefits of one having an S&C coach or personal trainer. Interestingly, a number of participants described the system either as a "virtual trainer" or a middle ground to having no personal trainer:

> *The app almost acts as a person telling you you're doing it wrong. That's how I felt.* [Beginner gym-goer]
>
> *I wouldn't get a personal trainer but this system could be a good middle ground.* [Beginner gym-goer]
>
> *If you don't want to hire a coach, as coaches are a lot of money then it will give you a pretty good overview of the kind of stuff you have to do.* [S&C coach]

There were no subthemes that emerged perceiving negative aspects to system use. However, one experienced gym-goer did suggest that use of the system could distract from focusing on their exercise technique. They stated the following:

> *It wasn't necessarily confusing but I did think I might be paying more attention to the app than my own form.* [Experienced gym-goer]

Subjective Quality

The subjective quality portion of the uMARS showed that, when available, 9 participants thought they would use this system 10 to 50 times over the next year, and 5 participants thought they would use it greater than 50 times in the next year. These 5 participants were all beginner gym-goers. All participants would recommend the system to people who might benefit from it. The median star rating from all 15 participants was 4 out of 5 (IQR: 3-4). The interview data reflected these quantitative ratings. All participants said they "liked" the system or thought it was "good." More detailed statements included:

> *Overall, I was very impressed with the app. I have to say, very impressed.* [Beginner gym-goer]
>
> *I've been going to the gym for whatever amount of years and I'd still use something like this if it can tell me which reps are good and which reps are bad.* [Experienced gym-goer]
>
> *In terms of something to use during a session, I think it would be great.* [S&C coach]

In terms of aspects of quality to improve on, 2 S&C coaches felt the feedback was perhaps a little basic and could be more detailed:

> *Not that the technology it involves is, but in terms of how much information you could actually access it was quite basic.* [S&C coach]

One experienced gym-goer also expressed that without more feedback, they might stop using the system once they had perfected their technique:

> *I think the limit to the app is once you have the motion down, you're less likely to keep using it.* [Experienced gym-goer]

Overall, all users subjective rating of system quality was positive. However, all participants had suggested improvements for future iterations of the app, which emerged as a theme during qualitative analysis and will be discussed in the upcoming subsection.

Future Changes

The most popular suggestion for future changes to the system was to add more exercises that can be tracked. Beginners stressed the need for this, saying things such as:

> *I would like it if there were more exercises within the app as standard gym session would generally involve more exercises.* [Beginner gym-goer]
>
> *I think just add more exercises. Keep developing it as it's just a great idea.* [Beginner gym-goer]
>
> *Maybe add some other type of movement that people do, I don't know how well it transfers to upper body movements but certainly bench press is something that people always tend to need help with when they first go in to a gym.* [S&C coach]
>
> *I guess just add more exercises. So then it would cover more things, because I guess there is a wide range of exercises that people do when they go to the gym and they can all be done with poor form if you don't know what you are doing.* [S&C coach]

Experienced gym-goers and S&C coaches regularly suggested the need for more exercises to be tracked by the system. However, because of their experience and knowledge, they also suggested the types of exercises that would require technique classification and suggested that for many exercises the system would only need to automatically count repetition and sets. There was a general consensus among the experienced gym-goers and S&C coaches that upper body compound exercises (eg, bench press, overhead press, and barbell row) were the additional exercises that should be incorporated to the system, including technique classification. There was also a shared opinion that users should be able to add any exercise they complete to the system to be logged automatically. However, it was suggested that noncompound, secondary exercises may not require technique classification, as they are associated with a lower injury risk. It was also said that Olympic weightlifting moves should not be added to the battery of assessed movements, as they would be too dangerous to learn via an app. Two statements that summarize the cohort's general opinion are as follows:

> *Because, it's in an app; I would say prioritize...you could have compound exercises like a bench press or an OHP (over-head press) but like you can't teach like a jerk or a clean so just compound or isolation movements as there is less that can go wrong with those kind of things.* [S&C coach]
>
> *In terms of other exercises; again I suppose I like the idea that it would manly be your key lifts. In terms of adding loads of other exercises, I don't know if it would be necessary. The ones we would mainly cover in terms of injury risk are your squat, your back squat, your deadlift etc. So yeah, in terms of that I'd keep it to key lifts.* [S&C coach]

Feedback was another prominent subtheme that emerged in the area of future changes. Two key things were suggested

recurrently: (1) providing longitudinal feedback or progress over time and (2) more detailed feedback on the completed exercise repetitions. With regards to providing longitudinal feedback, one beginner suggested the following:

> *I'd be really interested to use it regularly and see if I look back over weeks am I seeing progress?...(Comparison to SleepTracker) So I'd like to see something similar in this where you could link your exercise quality to your habits and progress.* [Beginner gym-goer]

This type of longitudinal feedback reflected what most users of the system would also desire. With regards to receiving more specific feedback on their completed exercise, a diverse range of suggestions were made:

> *Then after the end of your sets, if it tells you like an estimate of your maximum and could count your rest times.* [Beginner gym-goer]
>
> *It would be really interesting to actually see the angles.* [Experienced gym-goer]
>
> *A drop down with exactly what reps are good and bad would be useful.* [Experienced gym-goer]
>
> *We'd be quite keen on muscle fiber recruitment during an exercise. I'm not sure if the sensors can pick up on it.* [S&C coach]
>
> *Tempo—that would be a big one for us.* [S&C coach]

The most frequently reported request, however, was to receive feedback on the exact mistake one was making when exercising, as opposed to whether a repetition was simply "good" or "bad":

> *Maybe, for example, "your back was too arched" or "not arched enough," or the angle of your legs, how far down you should be going etc.* [Experienced gym-goer]

When participants mentioned this, they were informed of other work from the authors that uses multiple IMUs to classify the exact deviation one makes while completing the exercises [14,15,39]. They were then asked if they would rather prefer a multi-IMU system that may be more expensive than the evaluated single IMU system if it could identify specific exercise mistakes. Opinions were mixed on this, with 2 participants stating it "would depend on cost." However, one beginner, one experienced gym-goer, and one S&C coach did suggest they would actually prefer a multi-IMU system which had such capability:

> *I do actually think more sensors would be cool but I think, I think that because I'm a bit of a nerd with stuff, so I'm like more sensors, that's cool; more accurate data etc. I think for the people you may actually be selling this app to, one sensor is actually nearly too much.* [Beginner gym-goer]
>
> *I think I'd like more sensors and feedback. Wearing sensors doesn't put me off.* [Experienced gym-goer]
>
> *I suppose, because I'm dealing with high level athletes, I would prefer to have more sensors to get more information.* [S&C coach]

No other subthemes regarding future changes were found; however, one S&C coach did suggest a team version of the system, where multiple users could connect to a coach's tablet app. This would allow them to focus their time and attention to the team members who require it the most. They also suggested the straps and IMU should be improved to be more appealing in such a setting.

Discussion

Principal Findings

To the authors' knowledge, this study is the first to apply a mixed-methods approach to evaluating a wearable sensor-based exercise biofeedback system. In particular, this study is a first look at users' perceptions of such systems and their potential benefits. Therefore, in addition to providing information on the usability, functionality, and perceived impact of *Formulift*, the presented results also offer a number of end-user insights that can be leveraged to inform the development of future exercise biofeedback systems.

Usability and Functionality

The results demonstrated a good to excellent overall level of system usability. Participants highlighted that the *Formulift* system was easy to set up and intuitive in nature, particularly in relation to the ease at which they could complete tasks. This shows great promise for the uptake of wearable sensor-based exercise biofeedback apps within beginner and experienced gym-goer populations. However, analysis of the data also uncovered a number of specific usability issues that will be amended in future iterations of the app. For instance, the app should incorporate more status indicators, for example, the appearance of a loading screen while exercise data are being analyzed through the "Analyse my set" function. This addition would signal to the user that an action is taking place, thus reducing the user's uncertainty related to the completion of the task. A clearer method for navigating backwards in the app should also be added. Such changes should minimize confusion for system users and enable a more enjoyable and efficient user experience.

This study has also demonstrated that the *Formulift* system is functional. The key desired functions of *Formulift* are to accurately detect and count repetitions of the exercises under study, determine if each repetition was completed with "acceptable" or "aberrant" technique, and provide the user with interpretable feedback on their completed exercise. The combination of both qualitative and quantitative results shows that the system was indeed functional in these three areas. This was not withstanding a number of functionality bugs that were found during the study. The most significant bug was the real-time repetition counting algorithm lagging at the start of some user's sets. This must be rectified for future iterations of the system. It is essential that the real-time repetition counting functions correctly because if it does not, it may distract the user from completing their exercise properly.

Perceived Impact

One of the most important findings of this study is that the range of different system users (beginner gym-goers, experienced

gym-goers, and S&C coaches) reported several benefits to using the *Formulift* system. Most importantly, all users felt that the system would improve their technique as they exercise. This is a central finding as prior work has simply shown the ability of IMU-based exercise classification systems to detect "acceptable" and "aberrant" technique but has not determined if users would find feedback of this kind beneficial [12,15,17,39-41]. Interestingly, users also highlighted the system's positive psychological benefits with regards to improved levels of focus, motivation, and confidence while exercising. These perceived benefits are in line with desired aims for such systems as outlined in prior research [12,15,17,39-41]. Although these benefits are well reported aims for many biofeedback systems [42], the literature currently lacked end-user validation. Further study is required to objectively validate these perceived benefits. Despite no negative impacts of *Formulift*'s perceived impact emerging as subthemes, one participant did point out that the phone's position during exercise (ie, in the user's hand or on the floor in front of them) may be distracting from proper technique. This matter is not yet fully understood, and future iterations of the system should factor in how the phone is positioned during exercise to maximize the system's benefit to the user.

Subjective Quality

En masse, *Formulift* was well received by system users. The uMARS results showed the app had a median star rating of 4 out of 5. This shows that users thought the system was good but could also be improved. This feeling was backed up during participants' interviews. Although suggested improvements to the system will later be discussed, it is an important finding of this study that system users did like *Formulift*. Wearable sensor-based exercise biofeedback systems are a very new technology, and little is yet known about how users feel about using them. Therefore, it is encouraging to developers of such systems that the participants of this study gave predominantly positive feedback on the system.

Future Work

The systematic evaluation of the *Formulift* system uncovered a variety of suggested changes for future system iterations. Future work will endeavor to incorporate these changes in to the system or allow them to be customizable user preferences within the system, for example, turning on or off receiving feedback on tempo of movement. Additionally, future work will aim to establish users' perceptions of the system following multiple uses over a period of time. Such work would also grant the opportunity to (1) Objectively measure the impact of the system for users in a more rigorous manner and (2) Investigate how these findings compare to the system's perceived impact found in this study.

Comparison With Prior Work

As stated in the introduction to this paper, a vast proportion of the published work pertaining to exercise analysis with IMUs regards the efficacy of the various sensor setups and data analysis techniques to assess different exercises [10-19]. However, while in the early development phase of such systems, it is also of key importance to understand their usability, functionality, subjective quality, and perceived impact from the end user perspective. Involving the user early and often in the design process can help identify previously unforeseen user experience issues that can then be rectified to help increase levels of user engagement which is a central determinant to overall user adoption prospects [43].

To the author's knowledge, there are no published evaluation studies of wearable sensor-based exercise biofeedback systems. Literature is available in associated fields, with a vast variety of mHealth and exergaming systems being evaluated over the past decade [24,44]. It may be inappropriate to compare the results evaluating *Formulift* to apps and systems in other subdisciplines in mHealth because of the different demographics of users and purposes of such systems. However, it is important to note that the methodological approach undertaken in this study is in line with current state-of-the-art recommendations in usability studies [24]. Involving three types of real system users (beginners, experienced, and S&C) and employing a mixed-methods approach to evaluate the system has maximized our understanding on users' perceptions of *Formulift* and will inform the design of future iterations. Although recent work such as that by Kotsantinidis et al who outlined the development and evaluation of "FitForAll": an Exergaming Platform Improving Physical Fitness and Life Quality of Senior Citizens [45] has shown great benefits to exercise biofeedback systems and demonstrated good SUS scores, a lack of qualitative assessment of the system limits the conclusions that can be drawn on the system's usability and user experience. It is the authors' contention that we can maximize the learnings on mHealth system evaluations through combining appropriate surveys and interviews. This approach can more effectively inform the iterative design process to make the systems as beneficial as possible for end users.

Limitations

There are a number of contextual factors that should be considered when reviewing this study. To begin with, all results presented in this study are based on the participants' first use of the system. Although this is likely appropriate for highlighting any usability and functionality problems with the system, it is possible that one's rating of the system's quality and impact could vary over time. It should also be stated that results regarding the "perceived impact" of the system, as determined by the uMARS and thematic analysis of interview transcripts, are solely users' opinions on the benefits of the system. More work is required to determine if the system objectively improves, for example, people's motivation, exercise adherence, and exercise technique. To achieve this, a longitudinal study will be required. It should also be noted that this study took place in an artificial gym environment within a biomechanics laboratory. It would be useful to complete the proposed longitudinal system evaluation with participants exercising in their "normal" gyms. This may help uncover additional usability or functionality issues that should be amended and other future changes which could improve the system. A further limitation of this study is the sample size used and the homogeneity of the sample. While a sample of 15 participants complies with usability testing standards, it may not guarantee that the sample is representative of the wider population when considering

quantitative results. This may be the case in particular for specific populations not captured in our sample, for example, elderly, overweight, or underweight. However, there was high consistency when triangulating quantitative and qualitative results, which suggests they are both of merit. It should also be noted that this paper was concerned with determining the usability, functionality, perceived impact, and subjective quality of the *Formulift* system, and as such, expansive detail has not been provided on the following topics: the exact data analysis pathways utilized within the *Formulift* app; the quantitative performance of the system during the evaluation (ie, system accuracy, sensitivity, and specificity); and the manner by which the experienced exercisers and S&C coaches completed deliberately "aberrant" movements. These aspects of the system evaluation experiment are detailed in a recent paper by the authors (MOR, TW, and BC) [46]. This associated paper also details a tablet app that aims to ameliorate the overhead in setting up personalized classifiers for each exerciser, that is, the following processes are streamlined: synchronizing video and IMU data collection, signal processing, data segmentation, data labeling of segmented videos by an exercise professional, feature computation, and classifier creation [46]. Limiting the work involved in creating personalized exercise classification systems for new system users may be a key factor for their uptake in to the real world.

Conclusions

In this study, we sought to evaluate the *Formulift* system, a new exercise biofeedback app that classifies technique and tracks repetitions completed of exercises. A mixed-methods approach was undertaken to quantitatively and qualitatively assess the system under a variety of distinct categories: usability, functionality, subjective quality, and perceived impact. The assessment of the system was completed by three types of real system users: beginner gym-goers, experienced gym-goers, and qualified and practicing S&C coaches. The usability of the system was determined to be very good following both quantitative and qualitative analysis. The system also functioned as desired with users reporting that the system accurately detected their repetitions in real time, classified their exercise quality, and gave them appropriate feedback. Users expressed that they liked the system and that it could aid their focus and technique while exercising. Additionally, it was found that the system could increase their motivation and confidence in completing exercise. These findings are the first of their kind and show great promise for wearable sensor-based exercise biofeedback systems. However, this study also found a great deal of potential improvements to the *Formulift* system (Future Work subsection). By implementing these changes, it is hoped that systems such as *Formulift* may become an affordable, user-friendly, and useful tool that will aid gym-goers to enhance their training and support S&C coaches in the monitoring of their athletes.

Acknowledgments

This project is partly funded by the Irish Research Council as part of a Postgraduate Enterprise Partnership Scheme with Shimmer (EPSPG/2013/574) and partly funded by Science Foundation Ireland (SFI/12/RC/2289). The authors would also like to thank UCD Sport for providing equipment which was used in this study and express their gratitude to Ersi Ni who was pertinent to the development of the *Formulift* app.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video highlighting the *Formulift* system features and typical use case.

[MP4 File (MP4 Video), 93MB - mhealth_v6i1e33_app1.mp4]

Multimedia Appendix 2

List of tasks presented to users for system evaluation session.

[PDF File (Adobe PDF File), 23KB - mhealth_v6i1e33_app2.pdf]

Multimedia Appendix 3

Interview guide used for qualitative assessment of the *Formulift* system.

[PDF File (Adobe PDF File), 26KB - mhealth_v6i1e33_app3.pdf]

Multimedia Appendix 4

Table of themes arising from thematic analysis of interview transcripts.

[PDF File (Adobe PDF File), 70KB - mhealth_v6i1e33_app4.pdf]

References

1. Ahtiainen JP, Pakarinen A, Alen M, Kraemer WJ, Häkkinen K. Muscle hypertrophy, hormonal adaptations and strength development during strength training in strength-trained and untrained men. Eur J Appl Physiol 2003 Aug;89(6):555-563. [doi: 10.1007/s00421-003-0833-3] [Medline: 12734759]
2. Kraemer WJ, Mazzetti SA, Nindl BC, Gotshalk LA, Volek JS, Bush JA, et al. Effect of resistance training on women's strength/power and occupational performances. Med Sci Sports Exerc 2001 Jun;33(6):1011-1025. [Medline: 11404668]
3. Frontera WR, Meredith CN, O'Reilly KP, Knuttgen HG, Evans WJ. Strength conditioning in older men: skeletal muscle hypertrophy and improved function. J Appl Physiol (1985) 1988 Mar;64(3):1038-1044. [doi: 10.1152/jappl.1988.64.3.1038] [Medline: 3366726]
4. Bassett SF. The assessment of patient adherence to physiotherapy rehabilitation. N Z J Physiother 2003;31(2):60-66 [FREE Full text]
5. Friedrich M, Cermak T, Maderbacher P. The effect of brochure use versus therapist teaching on patients performing therapeutic exercise and on changes in impairment status. Phys Ther 1996 Oct;76(10):1082-1088. [Medline: 8863761]
6. Hall M, Nielsen JH, Holsgaard-Larsen A, Nielsen DB, Creaby MW, Thorlund JB. Forward lunge knee biomechanics before and after partial meniscectomy. Knee 2015 Dec;22(6):506-509. [doi: 10.1016/j.knee.2015.03.005] [Medline: 25907262]
7. Burns A, Greene BR, McGrath MJ, O'Shea TJ, Kuris B, Ayer SM, et al. SHIMMERTM–a wireless sensor platform for noninvasive biomedical research. IEEE Sensors J 2010 Sep;10(9):1527-1534. [doi: 10.1109/JSEN.2010.2045498]
8. Madgwick SO, Harrison AJ, Vaidyanathan A. Estimation of IMU and MARG orientation using a gradient descent algorithm. IEEE Int Conf Rehabil Robot 2011;2011:5975346. [doi: 10.1109/ICORR.2011.5975346] [Medline: 22275550]
9. McGrath D, Greene BR, O'Donovan KJ, Caulfield B. Gyroscope-based assessment of temporal gait parameters during treadmill walking and running. Sports Eng 2012 Jul 3;15(4):207-213. [doi: 10.1007/s12283-012-0093-8]
10. Seeger C, Buchmann A, Van Laerhoven K. myHealthAssistant: a phone-based body sensor network that captures the wearer's exercises throughout the day. 2011 Presented at: 6th International Conference on Body Area Networks; November 7-8; Beijing p. 1-7. [doi: 10.4108/icst.bodynets.2011.247015]
11. Morris T, Saponas S, Guillory A, Kelner I. RecoFit: using a wearable sensor to find, recognize, and count repetitive exercises. 2014 Presented at: CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26-May 1; Toronto, Ontario, Canada p. 2225-3234. [doi: 10.1145/2556288.2557116]
12. Giggins OM, Sweeney KT, Caulfield B. Rehabilitation exercise assessment using inertial sensors: a cross-sectional analytical study. J Neuroeng Rehabil 2014 Nov 27;11:158 [FREE Full text] [doi: 10.1186/1743-0003-11-158] [Medline: 25431092]
13. Whelan DF, O'Reilly MA, Ward TE, Delahunt E, Caulfield B. Technology in rehabilitation: evaluating the single leg squat exercise with wearable inertial measurement units. Methods Inf Med 2017;56(2):88-94. [doi: 10.3414/ME16-02-0002] [Medline: 27782290]
14. O Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Technology in strength and conditioning: assessing bodyweight squat technique with wearable sensors. J Strength Cond Res 2017 Aug;31(8):2303-2312. [doi: 10.1519/JSC.0000000000001957] [Medline: 28731981]
15. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield B. Classification of lunge biomechanics with multiple and individual inertial measurement units. Sports Biomech 2017 May 19;16(3):342-360. [doi: 10.1080/14763141.2017.1314544] [Medline: 28523981]
16. Whelan D, O Reilly M, Ward TE, Delahunt E, Caulfield B. Evaluating performance of the single leg squat exercise with a single inertial measurement unit. 2015 Presented at: REHAB '15 Proceedings of the 3rd 2015 Workshop on ICTs for improving Patients Rehabilitation Research Techniques; October 1-2; Lisbon, Portugal p. 144-147. [doi: 10.1145/2838944.2838979]
17. Kianifar R, Lee A, Raina S, Kulic D. Classification of squat quality with inertial measurement units in the single leg squat mobility test. Conf Proc IEEE Eng Med Biol Soc 2016 Dec;2016:6273-6276. [doi: 10.1109/EMBC.2016.7592162] [Medline: 28269683]
18. Giggins O, Kelly D, Caulfield B. Evaluating rehabilitation exercise performance using a single inertial measurement unit. 2013 Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health); May 5-8; Venice, Italy. [doi: 10.4108/icst.pervasivehealth.2013.252061]
19. Johnston W, O Reilly M, Dolan K, Reid N, Coughlan GF, Caulfield B. Objective classification of dynamic balance using a single wearable sensor. 2016 Presented at: 4th International Congress of Sports Science Research and Technology Support (icSPORTS); November 7-9; Porto, Potugal p. 15-24. [doi: 10.5220/0006079400150024]
20. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield BM. Classification of deadlift biomechanics with wearable inertial measurement units. J Biomech 2017;58:155-161. [doi: 10.1016/j.jbiomech.2017.04.028] [Medline: 28545824]
21. Al Ayubi SU, Parmanto B, Branch R, Ding D. A persuasive and social mHealth application for physical activity: a usability and feasibility study. JMIR Mhealth Uhealth 2014;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
22. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]

23. Lodhia V, Karanja S, Lees S, Bastawrous A. Acceptability, usability, and views on deployment of Peek, a mobile phone mHealth intervention for eye care in Kenya: qualitative study. JMIR Mhealth Uhealth 2016 May 09;4(2):e30 [FREE Full text] [doi: 10.2196/mhealth.4746] [Medline: 27160779]
24. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
25. Fitzgerald D, Kelly D, Ward T, Markham C, Caulfield B. Usability evaluation of e-motion: a virtual rehabilitation system designed to demonstrate, instruct and monitor a therapeutic exercise programme. 2008 Presented at: Virtual Rehabilitation (IWVR); August 25-28; Vancouver, BC, Canada p. 144-149. [doi: 10.1109/ICVR.2008.4625151]
26. Fitzgerald D, Trakarnratanakul N, Dunne L, Smyth B, Caulfield B. Development and user evaluation of a virtual rehabilitation system for wobble board balance training. Conf Proc IEEE Eng Med Biol Soc 2008;2008:4194-4198. [doi: 10.1109/IEMBS.2008.4650134] [Medline: 19163637]
27. Abran A, Khelifi A, Suryn W, Seffah A. Usability meanings and interpretations in ISO standards. Softw Qual J 2003;11(4):325-338. [doi: 10.1023/A:1025869312943]
28. Turner CW, Lewis JR, Nielsen J. Determining usability test sample size. Int Encycl Ergon Hum Factors 2006;3(2):3084-3088.
29. Sauro J. Measuring usability with the System Usability Scale (SUS). Userfocus.co.uk 2016 [FREE Full text]
30. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
31. Ritchie J, Lewis J. Qualitative Research Practice: A Guide for Social Science Students and Researchers. New York: Sage Publications; 2003.
32. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
33. Heath H, Cowley S. Developing a grounded theory approach: a comparison of Glaser and Strauss. Int J Nurs Stud 2004 Feb;41(2):141-150. [doi: 10.1016/S0020-7489(03)00113-5]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101 [FREE Full text]
35. Saldana J. The Coding Manual for Qualitative Researchers. New York: Sage Publications; 2009.
36. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
37. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9):1408-1416 [FREE Full text]
38. Brooke J. 1986. SUS - A quick and dirty usability scale URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2018-01-24] [WebCite Cache ID 6wi4L6Mlp]
39. O'Reilly MA, Whelan DF, Ward TE, Delahunt E, Caulfield B. Technology in Strength and conditioning: tracking lower limb exercises with wearable sensors. J Strength Cond Res 2017;31(6):1726-1736. [doi: 10.1519/JSC.0000000000001852] [Medline: 28538326]
40. Pernek I, Kurillo G, Stiglic G, Bajcsy R. Recognizing the intensity of strength training exercises with wearable sensors. J Biomed Inform 2015 Dec;58:145-155 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.020] [Medline: 26453822]
41. Taylor PE, Almeida GJ, Hodgins JK, Kanade T. Multi-label classification for the analysis of human motion quality. Conf Proc IEEE Eng Med Biol Soc 2012;2012:2214-2218. [doi: 10.1109/EMBC.2012.6346402] [Medline: 23366363]
42. Giggins OM, Persson UM, Caulfield B. Biofeedback in rehabilitation. J Neuroeng Rehabil 2013 Jun 18;10:60 [FREE Full text] [doi: 10.1186/1743-0003-10-60] [Medline: 23777436]
43. Abras C, Maloney-lrichmar D, Preece J. User-centered design. In: Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks, California: Sage Publications; 2004:445-455.
44. Sween J, Wallington SF, Sheppard V, Taylor T, Llanos AA, Adams-Campbell LL. The role of exergaming in improving physical activity: a review. J Phys Act Health 2014 May;11(4):864-870 [FREE Full text] [doi: 10.1123/jpah.2011-0425] [Medline: 25078529]
45. Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD. Design, implementation, and wide pilot deployment of FitForAll: an easy to use exergaming platform improving physical fitness and life quality of senior citizens. IEEE J Biomed Health Inform 2016 Jan;20(1):189-200. [doi: 10.1109/JBHI.2014.2378814] [Medline: 26731797]
46. O'Reilly M, Duffin J, Ward T, Caulfield B. Mobile app to streamline the development of wearable sensor-based exercise biofeedback Systems: system development and evaluation. JMIR Rehabil Assist Technol 2017 Aug 21;4(2):e9 [FREE Full text] [doi: 10.2196/rehab.7259] [Medline: 28827210]

Abbreviations

IMU: inertial measurement unit
IQR: interquartile range
mHealth: mobile health
S&C: strength and conditioning

SUS: system usability scale
uHealth: ubiquitous health
uMARS: user version of the Mobile Application Rating Scale

Edited by G Eysenbach; submitted 30.05.17; peer-reviewed by D Ding, A Billis, P Antoniou, C Mouzakidis, G Kurillo; comments to author 01.08.17; revised version received 29.08.17; accepted 28.09.17; published 31.01.18.

Please cite as:
O'Reilly MA, Slevin P, Ward T, Caulfield B
A Wearable Sensor-Based Exercise Biofeedback System: Mixed Methods Evaluation of Formulift
JMIR Mhealth Uhealth 2018;6(1):e33
URL: http://mhealth.jmir.org/2018/1/e33/
doi:10.2196/mhealth.8115
PMID:29386171

Original Paper

Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study

Florence Jimoh[1], BSc, MRes, PhD; Elizabeth K Lund[2], BSc, MSc, PhD; Linda J Harvey[3], BSc, PhD; Catherine Frost[1], BSc, MSc; W James Lay[1]; Mark A Roe[3], BSc; Rachel Berry[3], BSc; Paul M Finglas[3], BSc

[1]Food Angels UK Ltd, Newmarket, United Kingdom

[2]Independent Consultant, Norwich, United Kingdom

[3]Quadram Institute Bioscience, Norwich, United Kingdom

Corresponding Author:
Paul M Finglas, BSc
Quadram Institute Bioscience
Norwich Research Park
Norwich, NR4 7UA
United Kingdom
Phone: 44 1603255318
Fax: 44 1603507723
Email: paul.finglas@quadram.ac.uk

Abstract

Background: There is increasing recognition that personalized approaches may be more effective in helping people establish healthier eating patterns and exercise more, and that this approach may be particularly effective in adolescents.

Objective: The objective of this study was to investigate the use of a smartphone app (FoodWiz2) in supporting healthy lifestyle choices in adolescence.

Methods: Participants (N=34: 11 male, 23 female) aged 16-19 years in full- or part-time education were recruited from sixth form colleges, schools, and other further education establishments in Norfolk and Suffolk, United Kingdom, between February and May 2015. Participants recorded food intake and exercise using a paper diary for 4-5 weeks and then used the app for the same duration. Initial nutrition education and general support were provided during the paper diary use, but the app included personalized messages sent in response to app activity. At the end of each study phase, participants completed an online questionnaire to describe their experience of using the paper diary and app.

Results: Record completion declined throughout the study, possibly affected by examination pressure. Food intake data showed increased fruit consumption and significantly reduced consumption of chocolate snacks (P=.01) and fizzy drinks (P=.002) among participants using the app. Questionnaire responses indicated that the app was generally preferred to the paper diary, in particular, the app was seen as less boring to use (P=.03) and more acceptable in social settings (P<.001).

Conclusions: This app-based approach has shown the potential for a more effective approach to improving adolescent diet and exercise levels.

(JMIR Mhealth Uhealth 2018;6(1):e17) doi:10.2196/mhealth.7702

KEYWORDS

adolescent; smartphone app; diet; exercise; food intake; mobile applications

Introduction

A wide range of modern technologies designed to support health and well-being of individuals and specific populations are becoming available and increasingly affordable. This study was designed to investigate the use of a smartphone app (*FoodWiz2*) in supporting healthy lifestyle choices in adolescence. Adolescence is characterized by a period of rapid growth and development; the pubertal increase in height and weight coincides with changes in body composition, such as increased muscle and bone mass and fat deposition in girls [1]. Indeed, energy and nutrient requirements are greater in adolescence than at any other time [2]. The poor quality of many adolescent diets is recognized to be an important issue in relation to a range of short- and long-term health outcomes [3-5]. Recent data from large, cross-sectional surveys indicate that adolescent diets do

not meet dietary guidelines in the United Kingdom (Lower Reference Nutrient Intakes) or in Europe (Food-Based Dietary Guidelines, such as the Food Guide Pyramid) [6-8]. In a representative sample of the UK population, adolescents (11-18 years) reported consuming less than the recommended or reference amount of fruits, vegetables, and fiber, whereas saturated fat and sugar were more than the dietary reference value. Indeed, nonmilk extrinsic sugars accounted for 15.3% of total energy of this group compared with a reference of 11% [9]. In Europe, adolescent diets were similarly low in fruits and vegetables, as well as dairy, but high in meat and high-fat products and sweets [6].

Despite the implementation of public health campaigns, adherence to dietary advice is poor across many European countries [10,11]. In particular, it has been reported that adolescent diets may already have an impact on several health parameters [3,12-15], and yet, it is still far from clear which are the most effective approaches to change in lifestyle [16,17]. An important aspect of healthy living is the maintenance of a healthy weight that should be achieved by balancing energy intake with energy expenditure, and results of a recent systematic review indicated that diet, physical activity, and behavioral change interventions are effective in reducing body mass index (BMI) in overweight and obese adolescents [18]. Physical activity is considered an important component of a healthy lifestyle and should be encouraged as much as adherence to a good diet [19,20]. There is also a significant body of evidence to suggest that the effects of being overweight can be ameliorated by exercising regularly [21,22]. Furthermore, diet can improve exercise performance [23]. In general, the children of more highly educated parents exercise more and spend less time in a passive activity, such as computer use or watching television or DVDs [24]. Encouraging exercise as a part of healthy living needs to be particularly targeted at lower socioeconomic groups, as these have been identified as being mainly at risk and hard to reach [25], and the use of modern technologies may well go some way toward achieving this aim [26,27].

Dietary intake can be measured by a range of approaches, such as 24-hour recall, food frequency questionnaires (either self-administered or by interview asking about relatively long-term eating habits based on food groups), food diaries, or duplicate diet measurement. The first 3 approaches rely on using food composition databases, which are continuously being improved for content and accuracy. Of the 3 approaches, arguably the most accurate for measuring current food intake is the food diary method, preferably recording weighed food intake for as many days as possible. The main drawback with this approach is that it is labor-intensive and most people will either modify their diet to save writing down small snacks, record inaccurately, or just give up on recording [28]. It is well documented that completing food diaries, in combination with dietary advice, is associated with weight loss [29], and so, this approach has been widely utilized in helping adults lose weight and improve long-term eating habits. The benefit of this approach over following more extreme diets is that individual nutrient intakes can be monitored, and users can be informed about how well their diet matches with the national nutritional recommendations.

There is an increasing recognition that a personalized approach to nutritional modification may be more effective in helping people establish healthy eating patterns and, combined with encouragement to exercise more, lead to the establishment of better lifestyle habits [30,31]. Several recent studies have investigated the effectiveness of Web-based and smartphone apps in improving adherence to dietary advice in adults [32-38]. For example, 8 out of 42 people completed the app intake diary every day (defined as recording more than 500 kcal per day) in a group of overweight and obese adults using an app called *MyMealMate* over a period of 6 months [33]. However, in 2 parallel groups using either Web-based or paper-based approaches, only 1 or 2 people achieved this. The greater adherence to diary completion in the app group was associated with a reported benefit in terms of ease of use and a greater weight loss. Furthermore, modern technologies are being developed to help teenagers with specific health problems, such as type 1 diabetes [39]. Thus, the use of these technologies in adolescents to support healthy lifestyle choices is considered a potentially useful next step, which has been explored in several new studies [40-42]. *FoodWiz2* has been designed as a part of a European funded study to integrate different technologies to deliver personalized dietary advice to support health and well-being. Unlike some other currently available apps, it includes information on macronutrient content of foods taken from the UK food composition tables [43], allowing nutrient content of fresh and unprocessed foods to be available to the user, while other apps base their analysis on industry databases that may include data from a range of sources, including data from other countries.

Despite the proliferation of health-related apps and their apparent potential in dietary interventions, there is currently limited evidence on the experience of using these self-monitoring tools, and how participants perceive the comparison between novel and traditional methods of dietary assessment. Previous research into smartphone apps, personal digital assistants (PDAs), or short messaging service (SMS) interventions have attempted to assess participant experience, usually through questionnaires, and report on domains such as user satisfaction, patterns of usage, engagement, reasons for like or dislike, helpfulness, and influence on self-efficacy [33,35,44-50]. Studies in adolescent groups have examined the barriers and facilitators of using smartphone apps to record diet [42], using mobile technology (photos, emails, and texts) to record diet [41], and using diet recording apps for weight control [40]. These studies have also explored the effect of using different training methods (face-to-face vs telephone) [41] and the use of technology alone versus technology and counseling [40].

The aim of this study was to assess the ease of use, acceptability, and perceived effectiveness of a smartphone app for the measurement of food intake and exercise in adolescents compared with more traditional paper-based approaches.

Methods

Study Design and Ethics

All participants were initially asked to record food intake and exercise using a paper-based diary and then to use a smartphone app to record food and exercise as frequently and as accurately as they found feasible for 4-5 weeks. A 2-phase study design was chosen over a parallel intervention study to avoid introducing potential confounding effects, such as social factors, familiarity with the technology, and individual academic achievement. The study design also avoided potential carry-over effects and disproportionate dropout from the app-to-paper group that may have been associated with a conventional randomized crossover design. Personal and professional contacts were used to involve schools and colleges in the design of the study and to ensure that the protocol was appropriate in an educational environment with minimal disruption to pupils. The study was not blinded or randomized, but each participant was allocated a code for data analysis purposes. The study was scientifically reviewed by the Human Research Governance Committee at the Institute of Food Research, Norwich. The study was conducted according to the guidelines laid down in the Declaration of Helsinki, and all procedures involving human subjects were approved by the Oxford C Ethics Committee managed by the UK Health Research Authority (144/SC/1268).

Participants and Recruitment

Participants aged between 16 and 19 years, still in full- or part-time education, were recruited from sixth form colleges, schools, and other further education establishments in Norfolk and Suffolk, UK, between February and May 2015. The study was advertised through email and phone calls to education centers, followed by an initial presentation explaining the study first to the head and other interested staff and then to the students. The eligibility criteria were absence of chronic illness or disease (such as serious asthma or diabetes), not pregnant, able to give informed consent, have parental support, able or willing to use a smartphone, able to complete paperwork even with assistance, no eating disorders, no involvement in other research projects or weight management program, and not being related to any member of the study team. An inclusion criterion of BMI not below the second centile line on the BMI chart [51] was chosen as a safety measure to ensure that participants were not likely to fall below the defined healthy BMI range during the study. Signed letters of approval to run the study on each premises were obtained from the relevant school or college head (or other designated responsible person) at all participating schools (*Gatekeeper* approval). Written informed consent was obtained from all participants, and parental consent was also obtained for participants aged less than 18 years at recruitment.

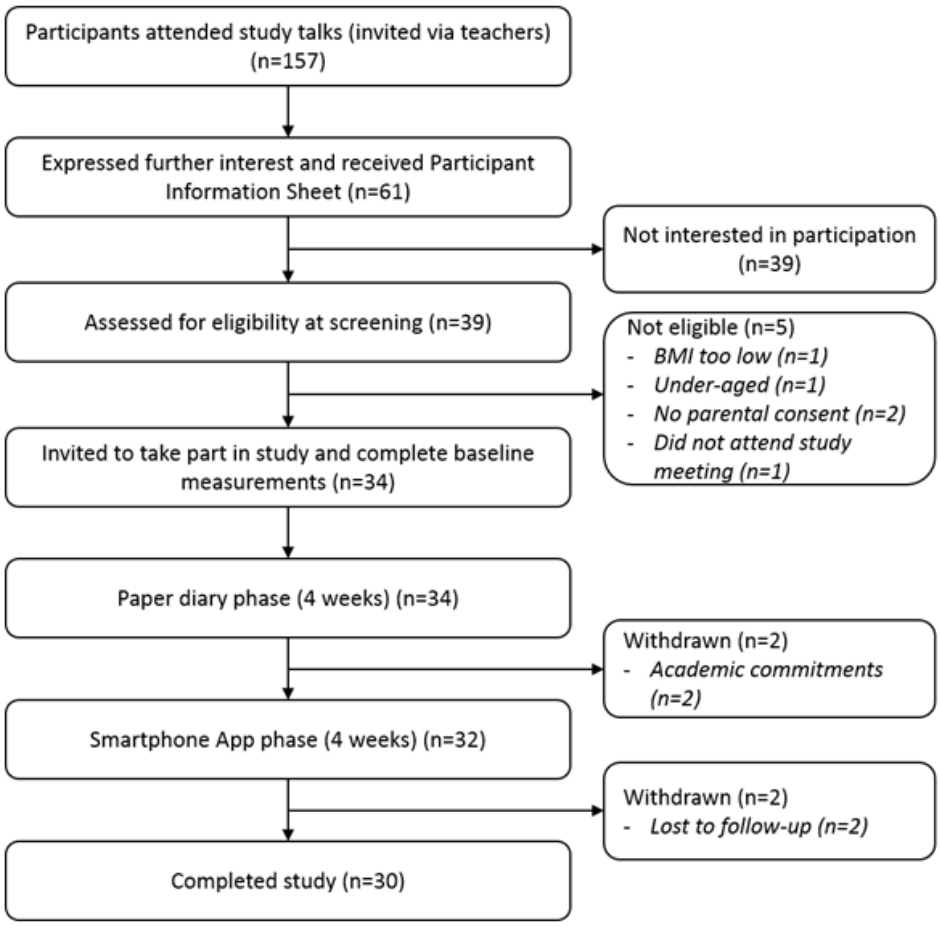
Eighteen schools in Norfolk and Suffolk were contacted via email, phone calls, or visit by the researcher. Five schools gave consent for the study talk to be delivered to their students. Recruitment for the study took place between February and May 2015. The flow of participants through the study is described in Figure 1. A total of 157 pupils attended study talks and 38 of these attended the screening session, where participants were assessed for eligibility to take part in the study. Of those screened, 4 were excluded for not meeting the inclusion criteria. Thirty-four adolescents consented and began the paper-diary phase. Two withdrew (the first person after the first week and the second person in the second week of the study) due to academic commitments. Thirty-two participants completed the first phase of the study and were given a smartphone preloaded with the *FoodWiz2* app. Two participants were lost to follow-up during this phase, with 30 completing the study. All 34 participants were, however, included in the analysis of the use of the paper diary, and the 32 that started using the app were included in its analysis.

Following consent, participants were screened based on the eligibility criteria. Weight was measured to the nearest 0.1 kg using a portable electronic scale (Salter Ultimate Accuracy, Tonbridge, Kent, UK); height was measured to the nearest millimeter using a portable stadiometer, and waist circumference was measured to the nearest millimeter using a body waist fitness caliper. Participants then completed a background information form, which provided more information about their diet, previous recording of diet or exercise, use of apps for similar purposes, and educational attainment. A score was calculated for the General Certificate of Secondary Education (GCSE) educational attainment based on a scale of 1-8 for grades G-A (including the higher level A* grade), and a cutoff of 25 (representing grade C in 5 subjects) was used to assess whether there was any difference in the interest in, or ability to, monitor diet among low and high academic achievers. The cutoff of 25 is used in the United Kingdom for assessing suitability at 16 years of age to progress to higher academic studies. Following screening, all participants received nutrition education sessions, covering the basics of a healthy diet, the importance of exercise, and different ways of achieving and maintaining a healthy nutrient intake and net energy intake based on government guidance.

Recording Diet and Exercise

Participants were initially provided with paper record sheets to record the day, date, time, type, and amount of all foods and drinks consumed; they could also record any recipe in the food record sheets. These food diaries were adapted from those previously used at The Institute of Food Research [52,53]. Participants also recorded the day, date, time type, duration, and intensity of any activity undertaken in the activity sheet. The paper diaries were collected and reviewed with participants every 2 weeks during a face-to-face meeting (10-30 min) between each participant and the researcher. If participants did not attend 2 consecutive meetings with the researchers and were not contactable, they were considered as no longer wishing to take part in the study (lost to follow-up).

Figure 1. Diagram showing the flow of participants through the study.



After a short interval in recording (1-2 weeks), coincident with a school or college holiday, a smartphone app developed by Food Angels UK Ltd, Newmarket, Suffolk, UK, was given to participants to use for the following 4 weeks (screenshots are available in Multimedia Appendix 1). *FoodWiz2* was developed from an earlier version (*FoodWiz*), which was designed to assist people suffering from allergies. Users of the first *FoodWiz* app can scan the barcode of a product to discover if it contains any allergens of interest (set up via their own personal profile). The more recent *FoodWiz2* app is like other food and exercise recording apps on the market and uses UK measures and databases rather than being US-based. The food composition data used in the app is taken from a commercially available source (BrandBank, Norwich, UK) combined with data for generic foods from the UK McCance and Widdowson's The Composition of Foods Integrated Dataset [43]. The app allows users to record dietary intake by searching and selecting from a list of foodstuffs and popular meal choices. The user can search by typing in the name of the food or by scanning in barcodes from products. The app also allows users to record intake of home-prepared recipes. A small set of scales, the size of a smartphone, was also provided to allow direct measurement of portion size weights using Bluetooth technology to link the scales to the app [54]. Physical activity could also be recorded using the App, allowing the user to receive instant feedback on their overall daily calorie balance or allowance based on set targets. *FoodWiz2* could also record and track weight and mood of participants. The researcher was able to see the information recorded by each participant using the app via a Web-based portal. Therefore, the researcher could monitor use and send each participant personalized feedback text messages for motivation and help. All participants were made aware that this would happen before the start of the study, both in person and in written form (Participant Information Sheet). Example feedback messages were approved during the ethical review process.

Ease of Use, Acceptability, and Perceived Effectiveness of the Recording Methods

At the end of each phase of the study, participants were asked to complete an online questionnaire (Multimedia Appendices 2 and 3). The questions were about the pattern of use, what it was like to use the tools, and the perceived impact on dietary and physical activity behavior. Examples of questions asked were, "To what extent do you agree with the following statements" e.g. on the usability ('was it time-consuming, disrupting, enjoyable, boring, convenient'), acceptability ('comfortable to use in social setting, would use in future') and perceived impact of each method of dietary assessment and exercise recording ('changed portion sizes of meals, ate more fruits and vegetables, increased awareness of physical activity, increased motivation to change physical activity'). An additional questionnaire section was completed at the end of the app phase (Multimedia Appendix 3) to assess the specific features of the app (eg, search function, feedback messages), and a final section compared the 2 recording methods. Questions were scored on

the Likert scale, but participants were also given the option of explaining some of their responses in more detail using a free-text approach. Free-text answers were independently reviewed by 2 researchers (FJ and PB), each identifying common topics before agreeing to the most important issues highlighted by participants. The statements included were agreed by both researchers. However, no computer-based systematic analysis of the free-text answers was undertaken due to the limited amount of data available.

Statistical Analysis

For the analysis of reported food intake using paper diaries or the app, descriptive measures were calculated to describe the sample by use of percentages, means, and standard deviations. Fruits and vegetable servings per day were calculated by dividing the total fruit and vegetable servings consumed by the total number of days that eating occasions were recorded using either the paper diary or App as described by Aflague et al [55]. The number of days a participant used the paper diary or App was set at the total number of days participants recorded food intake of ≥500 kcal.

Two analyses were conducted on questionnaire data because of missing responses: answers to individual questions were described as percentages of all data provided with a subsequent comparison of responses for just those who included their unique identity code, allowing paired analysis of the responses given. The significance of differences in the values of responses to the questionnaire after each phase of the study was evaluated using the sign test for nonparametric paired ordinal data. This test is more conservative than the more frequently used Wilcoxon signed rank test but more appropriate for this dataset where it is difficult to prove the difference between pairs is ordinally scaled [56].

Microsoft Excel 2010 for Windows was used to enter the data, and statistical analyses were performed using SPSS, Version 22 (SPSS Inc., Chicago, IL, USA). All tests were 2-tailed and $P<0.05$ was taken as indicating statistical significance.

Results

Participants and Recruitment

Participants (23 female, 11 male) aged 16 to 19 years were recruited in this study (Table 1). The mean participant BMI was 24 kg/m^2(SD 4) with 15% (5/34) classified as either clinically obese (BMI above the 98th centile based on their sex and age) or severely obese (BMI above 99.6th centile based on their sex and age). All participants remained weight-stable throughout the study with a median weight change of −0.1 kg (interquartile range [IQR]=−0.5 to −0.3, n=33) after the paper diary phase and a median change of 0.375 kg (IQR=−0.275 to −0.925, n=28) after the app phase compared with weight at the start of each phase (overall median study weight change=−0.05 kg; IQR=−1.87 to −0.488; n=30; *P*=.20). Of the 34 participants, 10 females and 1 male had previously recorded their diet or exercise. Six participants (4 female, 2 male) had previously used an app for recording food intake. Average GCSE point score was 40.6 and ranged from 0 to 85, with 58.8% (20/34) of participants scoring 25 and above. All except 3 participants who scored above 25 were attending sixth form schools. Participants who scored zero (n=6) were either registered to take their GCSEs (n=5) or had emigrated from outside of Europe (n=1), where the system of education was different.

Recording Diet and Exercise

The paper-diary phase was completed by 32 out of the 34 participants who started the intervention, whereas the subsequent App phase was completed by 30 out of the remaining 32 participants. Only 12% (4/34) of participants recorded on all the possible days (28 days in both phases). The mean number of study days completed (>500 kcal) for each 28-day phase is shown in Table 2. Use of the diet recording tool was highest in the paper diary group with a mean of 24 days (SD 6) completed compared with 17 days (SD 9) for the app (*P*=.002). There was no significant difference in completion rate between male and female participants, and no effect of educational attainment on completion of food records was identified at either phase.

The mean percentage completion for the paper diary was 86 (SD 10), whereas that for the app was 61 (SD 7), although it should be noted that most students were taking examinations during the App-based phase. Completion rate decreased gradually throughout the study period both during the paper diary phase and the app phase (Table 2), although use showed a trend toward an increase near the end of the app phase.

Reported food intakes were analyzed by food groups, only for those days where >500 kcal were recorded to allow for comparison between results from the paper diary and the app. In general, recorded food intake was similar using either method; however, the reported consumption of chocolate snacks (*P*=.01) and fizzy drinks (*P*=.002) was significantly lower during the app phase than when paper diaries were used (Figure 2). It was noted by researchers that the quality of the data retrievable from the app was considerably better in terms of specific foods eaten and sources. For example, although a paper diary may just say “Chinese take away,” the app would prompt for a more specific description, for example, “Noodles and sweet and sour chicken.”

All participants recorded a range of exercises in the paper diaries, most frequently walking and cycling but also team sports, gym, dance, and housework. Similarly, all those still in the study used the app for this purpose on at least 1 occasion with 2 participants providing data for at least 26 days, in comparison with 8 participants using the paper diary. Interestingly, there was a significant correlation between the numbers of days completed in the paper diary and the app (*P*<.001), although significantly more days were completed in the paper diary phase with a median of 16 days as compared with 6 days (*P*<.001).

Table 1. Characteristics of participants. SD: standard deviation; BMI: body mass index; GCSE: General Certificate of Secondary Education.

Characteristics	Male (n=11)	Female (n=23)
Age (years), mean (SD)	16.8 (0.8)	17.1 (0.85)
Weight (kg), mean (SD)	75.6 (12.76)	66.6 (12.59)
Height (m), mean (SD)	1.8 (0.08)	1.7 (0.06)
Waist circumference (cm), mean (SD)	85.0 (12.88)	79.1 (10.66)
BMI (kg/m^2), mean (SD)	24.6 (4.21)	24.2 (4.64)
BMI classification[a] , n		
Normal (below the 91st centile)	7	17
Overweight (above 91st centile)	1	4
Very overweight or clinically obese (above 98th centile)	2	0
Severely obese (above 99.6th centile)	1	2
Vegetarian or vegan, n	0	2[b]
Special diet, n	0	1[c]
Allergies, n	0	1[d]
Supplements, n	3[e]	0
Previously recorded diet or exercise, n	1	10
Previous use of diet or exercise app, n	2	4
Educational attainment classification[f] , n		
Below 25 GCSE points	5	9
25 and above GCSE points	6	14

[a]BMI thresholds vary by sex and one-year increments in age. The age range covered is 2-20 years (Boys UK and Girls UK, Body mass index, 2-20 years [51]).

[b]A participant became vegetarian after 2 weeks in the study (decision independent of the study).

[c]Mild intolerance to wheat and dairy.

[d]Penicillin allergy.

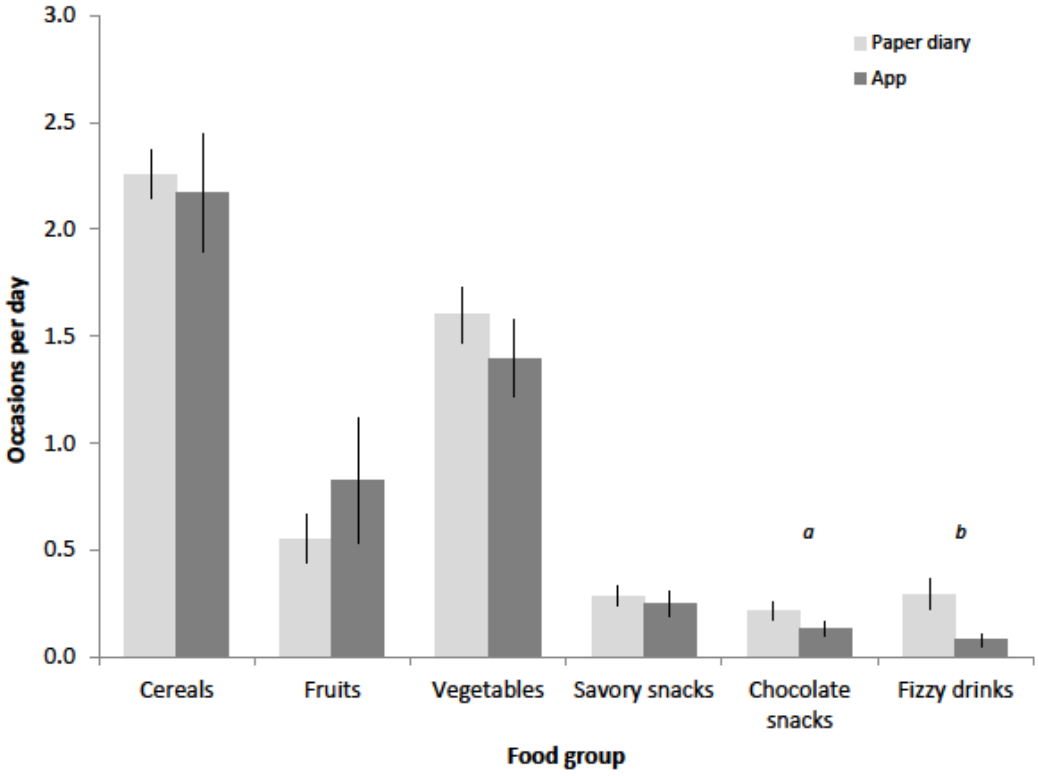
[e]Multivitamins and glucosamine phosphate (n=1); protein occasionally and vitamin tablets in winter (n=1); vitamin D, Branch chain amino acids, whey protein and creatine monohydrate (n=1).

[f]A score was calculated for GCSE based on a scale of 1-8 for grades G-A (including the higher level A* grade). The cutoff of 25 is equal to grade C in 5 subjects.

Table 2. Percentage completion of the study diary and App presented by weeks of the study. A completed day was regarded as a day with ≥500 kcal energy recorded [33]. SD: standard deviation.

Week	Percentage completion (SD)	Diet recording tool
1	96 (4)	Paper diet record
2	92 (7)	Paper diet record
3	82 (2)	Paper diet record
4	73 (7)	Paper diet record
5	66 (8)	App
6	61 (4)	App
7	54 (3)	App
8	64 (4)	App

Figure 2. Analysis of records per food group for the paper diary and app. Data are expressed as an average per day on which a total >500 kcal was recorded. Errors are expressed as standard error of the mean (SEM, n= 32). There is a significant difference between the use of paper diaries and the app in the recording of chocolate snacks and fizzy drinks (a: *P*=.012; b: *P*=.002).



Ease of Use, Acceptability, and Perceived Effectiveness of the Recording Methods

In most cases, no significant difference in response to Likert scale questions relating to the paper diary and the app were found. The free-text responses gave some additional insight into the issues raised (Multimedia Appendix 4). Although the number and range of responses were limited, they only gave a general indication of opinions. Review of the free-text answers given in the questionnaires suggested an overall preference for the *FoodWiz2* app, particularly in relation to enjoyment, convenience, recommendation to a friend, overall liking, and using again in the future, whereas the paper diary was considered time-consuming and boring. The main reasons given for preferring the app were focused on the topics of usability (eg, “Easy and fun to use”), accessibility (eg, “I can use it anywhere I want”), and the ability to track weight and calorie intake (eg, “Easy to watch calories”). Features of the app which were liked most were the smiley mood scale, bar code scanner, layout, and calorie counting.

The questions relating to patterns of use revealed that the reported level of use of the 2 approaches was similar, although the participants believed they used the paper diary for more days a week than the app (Question 1, Multimedia Appendix 2, *P*=.04), a result consistent with the actual data from the 2 recording methods. Participants reported that the paper diary was bulky and not easy to travel with, which often resulted in them recording what was eaten sometime after consumption. The problems associated with using a paper diary, perhaps only filling it when at home, would appear to have been more than outweighed by potential study fatigue and the issue of exam pressure during the app test phase (Multimedia Appendix 4).

Using both the paper diary and the app were considered time-consuming for different reasons. The paper diary involved having to manually write information and weigh foods, whereas the app had technical issues, for example, it worked slower than anticipated (Multimedia Appendix 4), partially explaining the lack of difference in response to the specific questions as to how time-consuming it was. Furthermore, although free-text answers suggested participants found the app more convenient, no significant difference was identified on the specific question (Section 1b-Q1, Multimedia Appendices 2 and 3). However, participants did report greater social acceptability for using the app (Figure 3). A higher proportion reported that they felt comfortable using their smartphone in social settings compared with using the paper diary (14/17 vs 3/21, *P*<.001). Although, when broken down into more specific social occasions, this effect was no longer significant and very few found using either approach difficult in front of their families. Furthermore, 48% (10/21) reported that they found the paper diary boring, whereas only 33% (5/15) found the app boring (*P*=.03, Figure 4).

Figure 3. Response to questionnaires in relation to how comfortable participants felt using the paper diary or app in different social scenarios. There is a significant difference between the use of paper diaries and the app in social settings (*P*<.001).

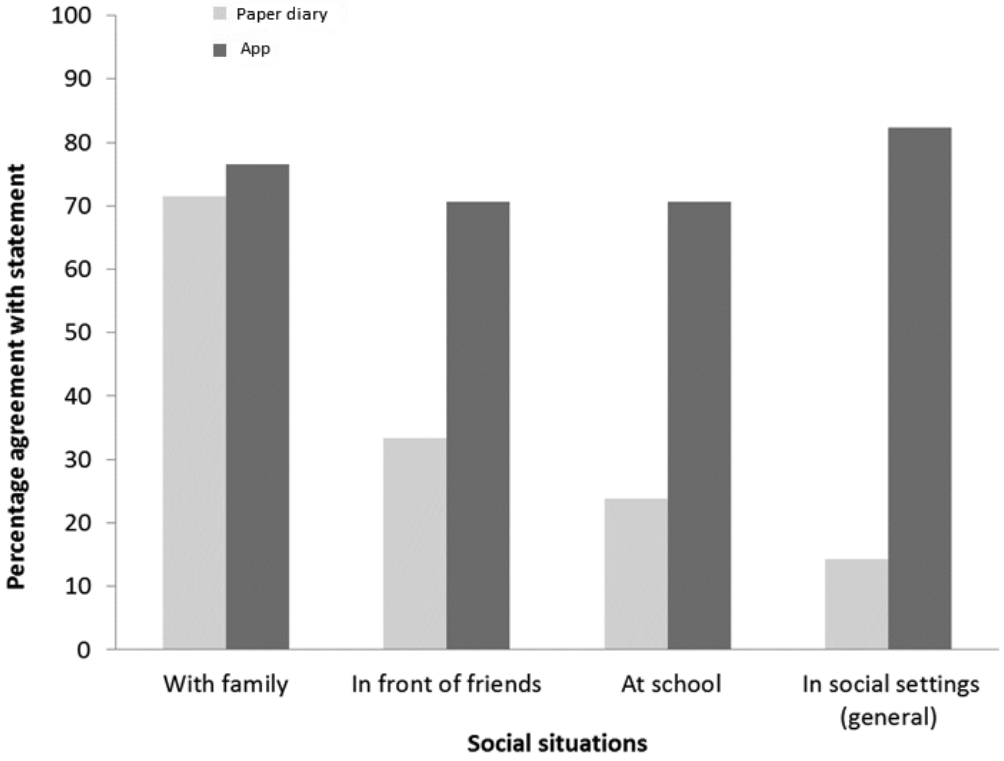
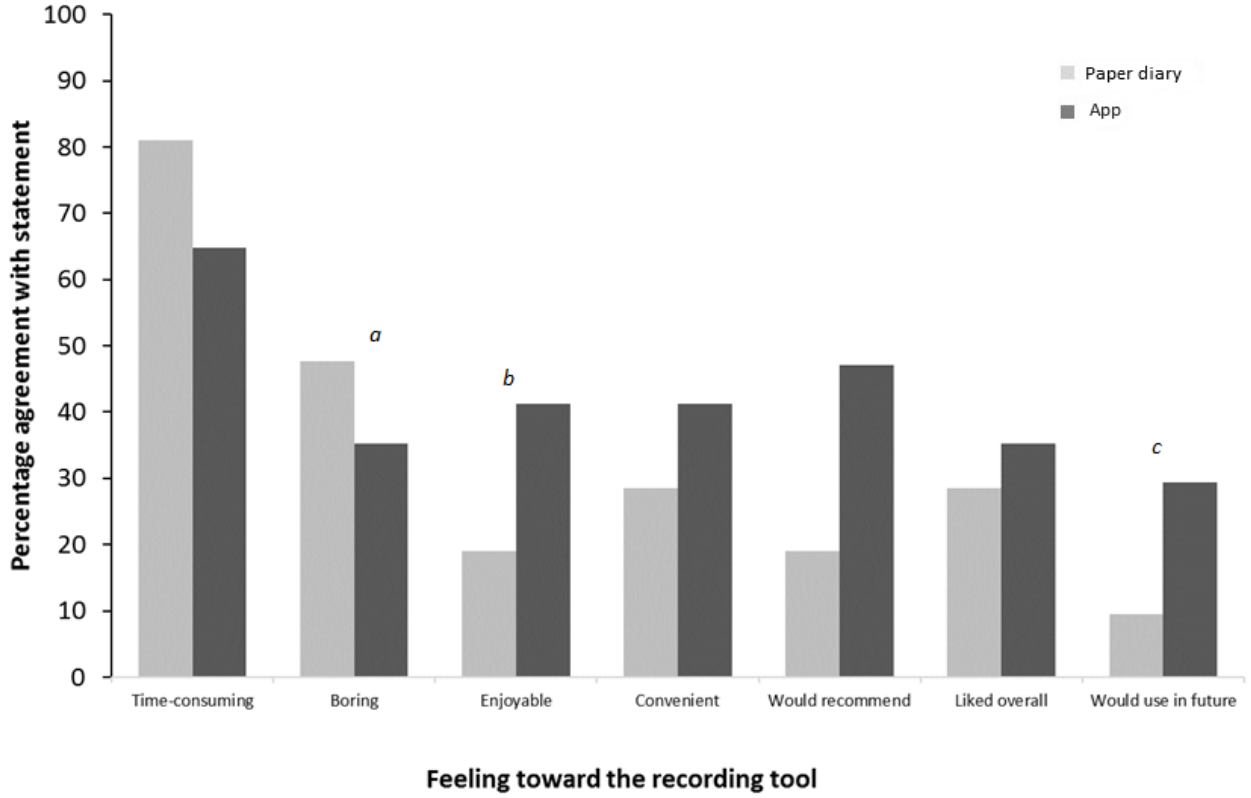


Figure 4. Response to questionnaires in relation to participants' general feelings towards using either the paper diary or app. For statements a, b, and c, there is a significant difference between the use of the paper diary and the app (a: *P*=.031; b: *P*=.04; c: *P*=.013).



Most participants reported that the use of both the paper diary and the app raised awareness of what they had been eating and how active they had been, and they felt that the app was more effective in this respect, which was again consistent with actual recorded data for chocolate snacks and fizzy drinks. However, there were no significant differences in perceived effectiveness between the 2 methods. When indicating the overall preference between the paper diary and the app, 5 out of 12 participants preferred the app, 3 out of 12 preferred the paper diary, and 4 out of 12 had no preference. Participants reported performing more aerobic and strength exercises as well as trying new activities; however, this was not significant, and indeed, the App appeared to significantly reduce confidence to do more activities. This reduction may be explained by the comments about the lack of options in the set list for exercise on the app (Multimedia Appendix 4). Specific questions addressing the quality of the app highlighted the value of being able to search for foods, but potential barriers to the use of the smartphone app were also identified. Feedback from the questionnaire showed that participants needed support with downloading and setting up *FoodWiz2* as it was not available via the usual app download routes (eg, Android Store). Occasionally, participants encountered difficulty finding nongeneric food products, which may have been a limitation of the search function. Suggested improvements to the app included healthier food recommendations, recipe ideas, more physical activity options, and help with portion sizes. The problems encountered with the app were mainly the speed of loading searches and ease of finding specific food items.

Discussion

Principal Findings

Adolescence is a key point in life in establishing long-term eating patterns as young people move forward into adult life with growing independence in food choice. Results of this study demonstrated that the app is a potentially feasible method of recording diet and physical activity in adolescents. Previous similar studies using smartphone apps with adults have focused on specific applications such as weight loss or caloric balance [33,49,57], including an intervention to encourage more attentive eating [58], as well as for other clinical issues such as pain management [59,60].

Strengths and Limitations of the Study

This study had several strengths including the fact that it focused on encouraging overall dietary improvement rather than calorie control in adolescents. This was considered a key issue, as some young people are very sensitive about their weight and may be particularly at risk of developing eating disorders such as anorexia or bulimia [61]. *FoodWiz2* included information from the UK food composition tables, so it was more comprehensive than those using solely commercial brand data sources that do not include products that are not prepacked, for example, fresh fruit, vegetables, and meat. The study also compared the new app technology with the standard paper diary, and participants were able to give feedback on their experience of using either recording method by answering both closed and open questions. The data were considered in the light of recorded food intake data as well as comparing intake data from the app and the paper-based diary. The trial retention in this study was higher (94% [30/32] of participants in the app phase) than that reported by Carter et al [57] testing a different app with women aged 35 years (SD 9). An equal number of participants were lost from the paper diary phase and the *FoodWiz2* app phase and were all linked with academic commitments. Participants were either in full-time or part-time education and so this was not unexpected. The study was subject to several challenges. Recruitment was lower than planned, such that only 57% (34/60) of the original recruitment target was reached. Direct contact with schools was the most successful recruitment strategy, but delays in starting arose as a result of one large potential source of participants withdrawing their support just as recruitment was about to start. This delay meant that the study ran over the normal school examination period in the United Kingdom from May to June. The frequency of use of the interventions was significantly higher in the paper diary compared with the app group (P=.002). This was unexpected and is not consistent with what has been previously reported [33,57,62]. This is likely to be due to two factors. First, the participants completed 4 weeks of recording using the paper diary before using the app for the same period and completion consistently dropped over this period, suggesting a fatigue effect. The use of the paper diary declined from the first week, but the use of the app increased slightly in the fourth and final week after exams had finished. The decline in the frequency of use of self-monitoring devices has been previously reported [57,63] in 2 studies comparing adherence to completion of dietary records using smartphone, website, or paper diary in randomized controlled trials. Second, review of the free-text answers highlighted that many participants were under pressure from academic commitments, especially those that had to prepare and appear for their advanced level examinations during the study, which impacted their ability to use the app fully. The comments by the participants that the paper diary was bulky and not easy to travel with resulting in them not recording what was eaten until sometime after consumption is likely to introduce inaccuracy in recording [49,64], although this was not detectable in this study. It is interesting to note that there was no significant difference between males and females or effect of educational attainment on the numbers of days on which food and exercise was recorded, although it should be noted this was a small self-selected subpopulation of all potential participants. The observed similarity in responses between male and female participants contrasts with the results reported in a recent study from the United States with children aged 3-10 years [55].

Effects of Interventions

Free-text responses in the questionnaires indicated that using either the paper diary or the *FoodWiz2* app raised participants' awareness of what they were eating and their level of activity. This also seemed to impact participants' view of their choice of food and the level of activity, particularly when participants used the app. This result is consistent with previous studies, where the recording of diet or activity has led to increased awareness and, in some instances, change in behavior [65-67]. Furthermore, the questionnaire data in this study suggesting that participants believed they had modified their intake of unhealthy foods, albeit not statistically significant, was reflected

in the results of the quantitative analysis of food entries in both the paper diary and the app, where significant reductions in chocolate snacks and fizzy drinks were reported using the app. In terms of acceptability and ease of use, participants were more comfortable using the app in different settings, especially in school and in social settings. These results are similar to those of previous studies [57,63]. In these previous studies, a significantly higher proportion of participants reported that the smartphone and website records were convenient to use compared with the paper-based record.

Conclusions

Our results indicate that, in general, participants preferred the use of a smartphone app to the more traditional paper diary, although some technical issues need to be addressed. In particular, participants found it more comfortable to use the app in social settings. They perceived that the use of the app had more impact on their dietary intake as well as physical activity, compared with the paper diary. Analysis of data from the recorded food intake also showed significantly reduced consumption of chocolate snacks and fizzy drinks among participants when they used the app to record their food intake compared with using the paper diary. The use of mobile technology shows great promise for reducing the burden of self-monitoring lifestyle in this age group, but future apps need to be more sophisticated than the one used for this study. Finally, the relative ease of data extraction for the app compared with coding food diaries and the quality of detail provided mean similar tools show great promise for research purposes.

Acknowledgments

The authors would like to thank the participants who recorded their diet and activities using the paper diary and the FoodWiz2 app, and head teachers, heads of sixth forms, and staff in the schools and colleges where the study was completed. The authors also thank Pippa Belderson, Department of Social Work, School of Health Sciences, University of East Anglia, Norwich, UK, for advice on the design of the study questionnaire and the analysis of the free-text responses.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613783 and from the Biotechnology and Biological Sciences Research Council Core Strategic Grant No. BB/J004510/1. The views expressed are those of the authors and not necessarily those of the European Union's Seventh Framework Programme for research, technological development and demonstration.

Conflicts of Interest

JL is the owner of Food Angels UK Ltd, which developed the FoodWiz2 app. FJ and CF were employed by Food Angels UK Ltd to complete the study. EL was contracted by Food Angels UK Ltd to assist with the study design and implementation. All the other authors declare no conflict of interest.

Multimedia Appendix 1

Screenshots from the FoodWiz2 app. A: the main menu; B: the home screen showing weight, daily targets, and mood; C: adding individual foods and portions.

[PDF File (Adobe PDF File), 184KB - mhealth_v6i1e17_app1.pdf]

Multimedia Appendix 2

The study questionnaire used following phase 1 (the paper diary).

[PDF File (Adobe PDF File), 409KB - mhealth_v6i1e17_app2.pdf]

Multimedia Appendix 3

The study questionnaire used following phase 2 (the App).

[PDF File (Adobe PDF File), 435KB - mhealth_v6i1e17_app3.pdf]

Multimedia Appendix 4

Free-text responses illustrating views expressed by participants responding to the questionnaire with free-text answers.

[PDF File (Adobe PDF File), 88KB - mhealth_v6i1e17_app4.pdf]

References

1. Jenkins S, Horner SD. Barriers that influence eating behaviors in adolescents. J Pediatr Nurs 2005 Aug;20(4):258-267. [doi: 10.1016/j.pedn.2005.02.014] [Medline: 16030505]

2. Department of Health. Dietary reference values for food energy and nutrients for the United Kingdom. Report of the panel on dietary reference values of the committee on medical aspects of food policy. London: HMSO; 1991.
3. Cuenca-García M, Ortega FB, Huybrechts I, Ruiz JR, González-Gross M, Ottevaere C, et al. Cardiorespiratory fitness and dietary intake in European adolescents: the healthy lifestyle in Europe by nutrition in adolescence study. Br J Nutr 2012 Jun;107(12):1850-1859. [doi: 10.1017/S0007114511005149] [Medline: 22117563]
4. Lytle LA, Kubik MY. Nutritional issues for adolescents. Best Pract Res Clin Endocrinol Metab 2003 Jun;17(2):177-189. [Medline: 12787546]
5. Murphy JM, Pagano ME, Nachmani J, Sperling P, Kane S, Kleinman RE. The relationship of school breakfast to psychosocial and academic functioning: cross-sectional and longitudinal observations in an inner-city school sample. Arch Pediatr Adolesc Med 1998 Sep;152(9):899-907. [Medline: 9743037]
6. Diethelm K, Jankovic N, Moreno LA, Huybrechts I, De Henauw S, De Vriendt T, HELENA Study Group. Food intake of European adolescents in the light of different food-based dietary guidelines: results of the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) Study. Public Health Nutr 2012 Mar;15(3):386-398. [doi: 10.1017/S1368980011001935] [Medline: 21936969]
7. Public Health England. 2014. National diet and nutrition survey: results from years 1-4 (combined) of the rolling programme (2008/9-2011/12) executive summary URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/310997/NDNS_Y1_to_4_UK_report_Executive_summary.pdf [accessed 2017-11-01] [WebCite Cache ID 6nREOOpfK]
8. Whitton C, Nicholson SK, Roberts C, Prynne CJ, Pot GK, Olson A, et al. National diet and nutrition survey: UK food consumption and nutrient intakes from the first year of the rolling programme and comparisons with previous surveys. Br J Nutr 2011 Dec;106(12):1899-1914 [FREE Full text] [doi: 10.1017/S0007114511002340] [Medline: 21736781]
9. Department of Health, Food Standards Agency. Gov.UK. 2012. National diet and nutrition survey URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207708/NDNS-Y3-report_All-TEXT-docs-combined.pdf [WebCite Cache ID 6nRIKzL7h]
10. de AD, Guessous I, Vaucher J, Preisig M, Waeber G, Vollenweider P, et al. Low compliance with dietary recommendations for food intake among adults. Clin Nutr 2013 Oct;32(5):783-788. [doi: 10.1016/j.clnu.2012.11.022] [Medline: 23260749]
11. Thorsdottir I, Birgisdottir B, Kiely M, Martinez J, Bandarra N. Fish consumption among young overweight European adults and compliance to varying seafood content in four weight loss intervention diets. Public Health Nutr 2009 May;12(5):592-598. [doi: 10.1017/S136898000800253X] [Medline: 18561865]
12. Vyncke KE, Huybrechts I, Dallongeville J, Mouratidou T, Van Winckel A, Cuenca-García M, et al. Intake and serum profile of fatty acids are weakly correlated with global dietary quality in European adolescents. Nutrition 2013 Feb;29(2):411-9.e1. [doi: 10.1016/j.nut.2012.07.007] [Medline: 23312763]
13. Vyncke K, Cruz FE, Fajó-Pascual M, Cuenca-García M, De Keyzer W, Gonzalez-Gross M, et al. Validation of the diet quality index for adolescents by comparison with biomarkers, nutrient and food intakes: the HELENA study. Br J Nutr 2013 Jun;109(11):2067-2078. [doi: 10.1017/S000711451200414X] [Medline: 23110799]
14. Patterson E, Wärnberg J, Kearney J, Sjöström M. Sources of saturated fat and sucrose in the diets of Swedish children and adolescents in the European Youth Heart Study: strategies for improving intakes. Public Health Nutr 2010 Dec;13(12):1955-1964. [doi: 10.1017/S1368980010001266] [Medline: 20450530]
15. Lund EK, Lee-Finglas WE, Southon S, Gee JM, Johnson IT, Finglas PM, et al. Dietary fat intake and plasma lipid levels in adolescents. Eur J Clin Nutr 1992 Dec;46(12):857-864. [Medline: 1483414]
16. Flynn MA, McNeil DA, Maloff B, Mutasingwa D, Wu M, Ford C, et al. Reducing obesity and related chronic disease risk in children and youth: a synthesis of evidence with 'best practice' recommendations. Obes Rev 2006 Feb;7 Suppl 1:7-66. [doi: 10.1111/j.1467-789X.2006.00242.x] [Medline: 16371076]
17. Waters E, de Silva-Sanigorski A, Hall BJ, Brown T, Campbell KJ, Gao Y, et al. Interventions for preventing obesity in children. Cochrane Database Syst Rev 2011(12):CD001871. [doi: 10.1002/14651858.CD001871.pub3] [Medline: 22161367]
18. Al-Khudairy L, Loveman E, Colquitt JL, Mead E, Johnson RE, Fraser H, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. Cochrane Database Syst Rev 2017 Dec 22(6):CD012691. [doi: 10.1002/14651858.CD012691] [Medline: 28639320]
19. Song M, Carroll DD, Fulton JE. Meeting the 2008 physical activity guidelines for Americans among U.S. youth. Am J Prev Med 2013 Mar;44(3):216-222. [doi: 10.1016/j.amepre.2012.11.016] [Medline: 23415117]
20. Dobbins M, Husson H, DeCorby K, LaRocca RL. School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18. Cochrane Database Syst Rev 2013(2):CD007651. [doi: 10.1002/14651858.CD007651.pub2] [Medline: 23450577]
21. Roberts CK, Katiraie M, Croymans DM, Yang OO, Kelesidis T. Untrained young men have dysfunctional HDL compared with strength-trained men irrespective of body weight status. J Appl Physiol (1985) 2013 Oct 01;115(7):1043-1049 [FREE Full text] [doi: 10.1152/japplphysiol.00359.2013] [Medline: 23887902]
22. De Bacquer D, Dallongeville J, Heidrich J, Kotseva K, Reiner Z, Gaita D, et al. Management of overweight and obese patients with coronary heart disease across Europe. Eur J Cardiovasc Prev Rehabil 2010 Aug;17(4):447-454. [doi: 10.1097/HJR.0b013e328336a05f] [Medline: 20485181]

23. Desbrow B, McCormack J, Burke LM, Cox GR, Fallon K, Hislop M, et al. Sports dietitians Australia position statement: sports nutrition for the adolescent athlete. Int J Sport Nutr Exerc Metab 2014 Oct;24(5):570-584. [doi: 10.1123/ijsnem.2014-0031] [Medline: 24668620]
24. Finger JD, Mensink GB, Banzer W, Lampert T, Tylleskär T. Physical activity, aerobic fitness and parental socio-economic position among adolescents: the German health interview and examination survey for children and adolescents 2003-2006 (KiGGS). Int J Behav Nutr Phys Act 2014 Mar 22;11(1):43 [FREE Full text] [doi: 10.1186/1479-5868-11-43] [Medline: 24656205]
25. Roberts K, Cavill N, Hancock C, Rutter H. Activenorfolk. Oxford: Public Health England Obesity Knowledge and Intelligence; 2013. Social and economic inequalities associated with diet and physical activity, including possible explanations for these inequalities URL: http://www.activenorfolk.org/files/10495/vid-19253-social-and-economic-inequalities-in-diet-and-physical-activity-041113.pdf [accessed 2017-09-08] [WebCite Cache ID 6saPMGOED]
26. Lioret S, McNaughton SA, Cameron AJ, Crawford D, Campbell KJ, Cleland VJ, et al. Three-year change in diet quality and associated changes in BMI among schoolchildren living in socio-economically disadvantaged neighbourhoods. Br J Nutr 2014 Jul 28;112(2):260-268. [doi: 10.1017/S0007114514000749] [Medline: 24775601]
27. Staiano AE, Abraham AA, Calvert SL. Motivating effects of cooperative exergame play for overweight and obese adolescents. J Diabetes Sci Technol 2012 Jul 01;6(4):812-819 [FREE Full text] [doi: 10.1177/193229681200600412] [Medline: 22920807]
28. Gibson R. Measurement errors in dietary assessment. In: Principles of Nutritional Assessment, 2nd edition. New York: Oxford University Press; 2005.
29. Kong A, Beresford SA, Alfano CM, Foster-Schubert KE, Neuhouser ML, Johnson DB, et al. Self-monitoring and eating-related behaviors are associated with 12-month weight loss in postmenopausal overweight-to-obese women. J Acad Nutr Diet 2012 Sep;112(9):1428-1435 [FREE Full text] [doi: 10.1016/j.jand.2012.05.014] [Medline: 22795495]
30. Anderson AS, Caswell S, Wells M, Steele RJ, Macaskill S. "It makes you feel so full of life" LiveWell, a feasibility study of a personalised lifestyle programme for colorectal cancer survivors. Support Care Cancer 2010 Apr;18(4):409-415. [doi: 10.1007/s00520-009-0677-4] [Medline: 19554354]
31. Elley CR, Dean S, Kerse N. Physical activity promotion in general practice--patient attitudes. Aust Fam Physician 2007 Dec;36(12):1061-1064 [FREE Full text] [Medline: 18075637]
32. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
33. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
34. Hebden L, Balestracci K, McGeechan K, Denney-Wilson E, Harris M, Bauman A, et al. 'TXT2BFiT' a mobile phone-based healthy lifestyle program for preventing unhealthy weight gain in young adults: study protocol for a randomized controlled trial. Trials 2013;14:75 [FREE Full text] [doi: 10.1186/1745-6215-14-75] [Medline: 23506013]
35. Hebden L, Cook A, vander ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2014 Aug;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
36. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
37. Patrick K, Marshall SJ, Davila EP, Kolodziejczyk JK, Fowler JH, Calfas KJ, et al. Design and implementation of a randomized controlled social and mobile weight loss trial for young adults (project SMART). Contemp Clin Trials 2014 Jan;37(1):10-18 [FREE Full text] [doi: 10.1016/j.cct.2013.11.001] [Medline: 24215774]
38. Recio-Rodríguez JI, Martín-Cantera C, González-Viejo N, Gómez-Arranz A, Arietaleanizbeascoa MS, Schmolling-Guinovart Y, et al. Effectiveness of a smartphone application for improving healthy lifestyles, a randomized clinical trial (EVIDENT II): study protocol. BMC Public Health 2014;14:254 [FREE Full text] [doi: 10.1186/1471-2458-14-254] [Medline: 24628961]
39. Hertel NT, Vedel K, Rohde L, Olesen JB. Serious disease - serious game. Stud Health Technol Inform 2013;192:1166. [Medline: 23920940]
40. Jensen CD, Duncombe KM, Lott MA, Hunsaker SL, Duraccio KM, Woolford SJ. An evaluation of a smartphone-assisted behavioral weight control intervention for adolescents: pilot study. JMIR Mhealth Uhealth 2016 Aug 23;4(3):e102 [FREE Full text] [doi: 10.2196/mhealth.6034] [Medline: 27554704]
41. Segovia-Siapco G, Sabaté J. Using personal mobile phones to assess dietary intake in free-living adolescents: comparison of face-to-face versus telephone training. JMIR Mhealth Uhealth 2016;4(3):e91 [FREE Full text] [doi: 10.2196/mhealth.5418] [Medline: 27473291]
42. Svensson A, Magnusson M, Larsson C. Overcoming barriers: adolescents' experiences using a mobile phone dietary assessment app. JMIR Mhealth Uhealth 2016 Jul 29;4(3):e92 [FREE Full text] [doi: 10.2196/mhealth.5700] [Medline: 27473462]

43. UK Government.: Public Health England; 2015. Composition of foods integrated dataset (CoFID) URL: https://www.gov.uk/government/publications/composition-of-foods-integrated-dataset-cofid [WebCite Cache ID 6saOU8Ja3]
44. Gasser R, Brodbeck D, Degen M, Luthiger J, Wyss R, Reichlin S. Springer. 2006. Persuasiveness of a mobile lifestyle coaching application using social facilitation URL: https://link.springer.com/chapter/10.1007%2F11755494_5 [accessed 2017-12-11] [WebCite Cache ID 6vdT2Sd6C]
45. Gerber BS, Stolley MR, Thompson AL, Sharp LK, Fitzgibbon ML. Mobile phone text messaging to promote healthy behaviors and weight loss maintenance: a feasibility study. Health Informatics J 2009 Mar;15(1):17-25 [FREE Full text] [doi: 10.1177/1460458208099865] [Medline: 19218309]
46. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. Public Health Nutr 2009 Dec;12(12):2382-2391. [doi: 10.1017/S1368980009005230] [Medline: 19323865]
47. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
48. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]
49. Tsai CC, Lee G, Raab F, Norman G, Sohn T, Griswold W, et al. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. Mobile Netw Appl 2007;12(2-3):173-184.
50. Yon BA, Johnson RK, Harvey-Berino J, Gold BC, Howard AB. Personal digital assistants are comparable to traditional diaries for dietary self-monitoring during a weight loss program. J Behav Med 2007 Apr;30(2):165-175. [doi: 10.1007/s10865-006-9092-1] [Medline: 17216341]
51. Royal College of Paediatrics and Child Health. 2013. UK CPCM growth chart, girls and boys, 2-20 years URL: http://www.rcpch.ac.uk/system/files/protected/page/GIRLS%20and%20BOYS%20BMI%20CHART.pdf [WebCite Cache ID 6nRJEvHUy]
52. Pot GK, Majsak-Newman G, Geelen A, Harvey LJ, Nagengast FM, Witteman BJ, FISHGASTRO Study Group. Fish consumption and markers of colorectal cancer risk: a multicenter randomized controlled trial. Am J Clin Nutr 2009 Aug;90(2):354-361 [FREE Full text] [doi: 10.3945/ajcn.2009.27630] [Medline: 19553301]
53. Heath AM, Roe MA, Oyston SL, Fairweather-Tait SJ. Meal-based intake assessment tool: relative validity when determining dietary intake of Fe and Zn and selected absorption modifiers in UK men. Br J Nutr 2005 Mar;93(3):403-416. [Medline: 15877881]
54. Computer Systems: Jožef Stefan Institute. 2016. Wireless pocket-size kitchen scale URL: http://cs.ijs.si/applications/btscale/ [WebCite Cache ID 6nSgdobhC]
55. Aflague TF, Boushey CJ, Guerrero RT, Ahmad Z, Kerr DA, Delp EJ. Feasibility and use of the mobile food record for capturing eating occasions among children ages 3-10 years in Guam. Nutrients 2015 Jun;7(6):4403-4415 [FREE Full text] [doi: 10.3390/nu7064403] [Medline: 26043037]
56. Elvers G. Udayton. Using SPSS for ordinally scaled data: mann-whitney u, sign test, and wilcoxon tests URL: http://academic.udayton.edu/gregelvers/psy216/spss/ordinaldata.htm [accessed 2017-02-08] [WebCite Cache ID 6sPkjq24h]
57. Carter MC, Burley VJ, Nykjaer C, Cade JE. 'My Meal Mate' (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
58. Robinson E, Higgs S, Daley AJ, Jolly K, Lycett D, Lewis A, et al. Development and feasibility testing of a smart phone based attentive eating intervention. BMC Public Health 2013;13:639 [FREE Full text] [doi: 10.1186/1471-2458-13-639] [Medline: 23837771]
59. Jibb LA, Stevens BJ, Nathan PC, Seto E, Cafazzo JA, Stinson JN. A smartphone-based pain management app for adolescents with cancer: establishing system requirements and a pain care algorithm based on literature review, interviews, and consensus. JMIR Res Protoc 2014;3(1):e15 [FREE Full text] [doi: 10.2196/resprot.3041] [Medline: 24646454]
60. de la Vega R, Roset R, Castarlenas E, Sánchez-Rodríguez E, Solé E, Miró J. Development and testing of painometer: a smartphone app to assess pain intensity. J Pain 2014 Oct;15(10):1001-1007. [doi: 10.1016/j.jpain.2014.04.009] [Medline: 24854065]
61. The McKnight Investigators. Risk factors for the onset of eating disorders in adolescent girls: results of the McKnight longitudinal risk factor study. Am J Psychiatry 2003 Feb;160(2):248-254. [doi: 10.1176/ajp.160.2.248] [Medline: 12562570]
62. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
63. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
64. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ 2002 May 18;324(7347):1193-1194 [FREE Full text] [Medline: 12016186]

65. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
66. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
67. Gilliland J, Sadler R, Clark A, O'Connor C, Milczarek M, Doherty S. Using a smartphone application to promote healthy dietary behaviours and local food consumption. Biomed Res Int 2015;2015:841368 [FREE Full text] [doi: 10.1155/2015/841368] [Medline: 26380298]

Abbreviations

BMI: body mass index
GCSE: General Certificate of Secondary Education
PDAs: personal digital assistants
SD: standard deviation
SMS: short messaging service

Edited by G Eysenbach; submitted 15.03.17; peer-reviewed by P Colombani, K Ng; comments to author 15.07.17; revised version received 10.08.17; accepted 30.10.17; published 15.01.18.

Please cite as:
Jimoh F, Lund EK, Harvey LJ, Frost C, Lay WJ, Roe MA, Berry R, Finglas PM
Comparing Diet and Exercise Monitoring Using Smartphone App and Paper Diary: A Two-Phase Intervention Study
JMIR Mhealth Uhealth 2018;6(1):e17
URL: http://mhealth.jmir.org/2018/1/e17/
doi:10.2196/mhealth.7702
PMID:29335239

Original Paper

Evaluating Machine Learning–Based Automated Personalized Daily Step Goals Delivered Through a Mobile Phone App: Randomized Controlled Trial

Mo Zhou[1], MS; Yoshimi Fukuoka[2,3], RN, PhD, FAAN; Yonatan Mintz[1], MS; Ken Goldberg[1,4], PhD; Philip Kaminsky[1], PhD; Elena Flowers[3,5], RN, PhD; Anil Aswani[1], PhD

[1]Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA, United States
[2]Institute for Health & Aging, School of Nursing, University of California, San Francisco, CA, United States
[3]Department of Physiological Nursing, School of Nursing, University of California, San Francisco, CA, United States
[4]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, United States
[5]Institute for Human Genetics, University of California, San Francisco, CA, United States

Corresponding Author:
Anil Aswani, PhD
Department of Industrial Engineering and Operations Research
University of California
4119 Etcheverry Hall
Berkeley, CA, 94720
United States
Phone: 1 510 664 9114
Email: aaswani@berkeley.edu

Abstract

Background: Growing evidence shows that fixed, nonpersonalized daily step goals can discourage individuals, resulting in unchanged or even reduced physical activity.

Objective: The aim of this randomized controlled trial (RCT) was to evaluate the efficacy of an automated mobile phone–based personalized and adaptive goal-setting intervention using machine learning as compared with an active control with steady daily step goals of 10,000.

Methods: In this 10-week RCT, 64 participants were recruited via email announcements and were required to attend an initial in-person session. The participants were randomized into either the intervention or active control group with a one-to-one ratio after a run-in period for data collection. A study-developed mobile phone app (which delivers daily step goals using push notifications and allows real-time physical activity monitoring) was installed on each participant's mobile phone, and participants were asked to keep their phone in a pocket throughout the entire day. Through the app, the intervention group received fully automated adaptively personalized daily step goals, and the control group received constant step goals of 10,000 steps per day. Daily step count was objectively measured by the study-developed mobile phone app.

Results: The mean (SD) age of participants was 41.1 (11.3) years, and 83% (53/64) of participants were female. The baseline demographics between the 2 groups were similar (P>.05). Participants in the intervention group (n=34) had a decrease in mean (SD) daily step count of 390 (490) steps between run-in and 10 weeks, compared with a decrease of 1350 (420) steps among control participants (n=30; P=.03). The net difference in daily steps between the groups was 960 steps (95% CI 90-1830 steps). Both groups had a decrease in daily step count between run-in and 10 weeks because interventions were also provided during run-in and no natural baseline was collected.

Conclusions: The results showed the short-term efficacy of this intervention, which should be formally evaluated in a full-scale RCT with a longer follow-up period.

Trial Registration: ClinicalTrials.gov: NCT02886871; https://clinicaltrials.gov/ct2/show/NCT02886871 (Archived by WebCite at http://www.webcitation.org/6wM1Be1Ng).

(JMIR Mhealth Uhealth 2018;6(1):e28) doi:10.2196/mhealth.9117

KEYWORDS

physical activity; cell phone; fitness tracker; clinical trial

Introduction

Physical Inactivity

Physical inactivity is the fourth leading risk factor for mortality, causing an estimated 3.2 million deaths worldwide [1]. It is associated with cardiovascular disease, certain types of cancer, type 2 diabetes, and depression [2-5]. Moderate- to vigorous-intensity physical activity, such as brisk walking or running, has significant health benefits across all age groups. The 2008 National Physical Activity Guideline for Americans recommends at least either 150 min of moderate-intensity physical activity or 75 min a week of vigorous-intensity physical activity for adults [6]. However, approximately half of American adults, particularly women and minorities, do not meet this physical activity guideline [7,8].

Mobile Health Interventions

Several lifestyle modification programs that promote physical activity have been demonstrated to be effective, but these programs are costly and labor-intensive because they require substantial in-person counseling [9-11]. To lower costs, researchers have conducted randomized controlled trials (RCTs) to investigate the feasibility of mobile health (mHealth) interventions (eg, mobile phone apps and digital pedometers) with reduced number of in-person counseling sessions [12-20]. Prior mHealth interventions implemented various goal-setting strategies to induce efforts, for example, to achieve and maintain 10,000 steps per day [21-26] or meet adaptively increasing step goals [14,27-29]. These studies demonstrated that mHealth interventions with goal setting can increase physical activity relative to baseline levels of activity.

Goal Setting

Goal setting is known to be an important factor for facilitating behavior change [30-32], and effective goal setting requires self-monitoring to better enable attainment of goals and increase self-efficacy [14,30,31,33]. There are three considerations regarding goal setting: (1) self-set goals versus assigned goals versus participatory goals, (2) adaptive goals versus fixed goals, and (3) personalized goals versus nonpersonalized goals. Despite the fact that self-set goals are of higher personal importance, a review of the goal-setting literature [30] reveals that assigned goals are more effective compared with self-set goals because self-set goals require regular input from participants, which is more difficult to maintain. Furthermore, more recent RCTs reveal that increases in physical activity through mobile-only programs with fixed, nonpersonalized physical activity goals are often substantially lower than increases in physical activity through programs that include adaptive goals [28,29,34,35] or personalized goal setting provided during in-person counseling [13,34,36-39]. For instance, one study [29] found that setting adaptive step goals resulted in an increase of 1130 more steps between baseline and 6 months, compared with setting fixed step goals of 10,000. Studies suspect that assigning nonpersonalized, fixed goals to all participants can lead to unrealistically high goals for some participants and unchallenging goals for other participants, which reduces goal-setting effectiveness [37,40]. Therefore, assigning adaptively personalized goals can be a favorable alternative to better induce efforts and increase physical activity [40-42]. Personalized, adaptive goal setting allows changing goals over time based on prior individual behavior. For example, future daily step goals can be assigned based on step totals from the previous days to ensure that the goals are challenging yet realistic for each individual. Two trials [28,29] used the same approach by combining financial incentives for meeting goals with an adaptive approach that set goals for the next day to be the 60th percentile of the steps taken in the past 10 days. Although this simple adaptive goal algorithm was modestly effective [28], a computer simulation study [43] for a weight loss intervention involving physical activity goal-setting and in-person counseling sessions found that a more sophisticated algorithm using statistics and machine learning to set goals by learning participants' responsiveness to goals could provide greater effectiveness (as compared with simple rules such as goal setting using a fixed percentile of steps taken in the past few days) in encouraging individuals to increase their physical activity and lose weight. In particular, the simulation showed that (when each participant received four counseling sessions) the more sophisticated machine learning algorithm would encourage almost half of the participants to have 5% or more body weight loss, whereas the use of goal setting using a fixed percentile of steps taken in the past few days would encourage only about one-quarter of the participants to have 5% or more body weight loss. Furthermore, previous studies have found that financial incentives may be effective during the intervention period, but in the maintenance period, participants are more likely to not adhere when no financial incentive is given [44-46].

Study Purpose

The purpose of our study was to test a sophisticated algorithm for personalized, adaptive goal setting that uses statistics and machine learning [43,47], and specifically to examine its efficacy in a fully automated mobile phone–based intervention with no in-person contact or counseling sessions during the trial. It is important to note that goal setting is only one component of a behavior change intervention, and our study is designed to isolate the impact of goal setting from other components to evaluate the efficacy of goal setting alone. We developed an automated mobile phone–based iPhone operating system (iOS, Apple Inc) app named CalFit, which sets personalized, adaptive step goals using the behavioral analytics algorithm (BAA) [43,47], and conducted an RCT (called Cal Fitness) using this mobile phone app in the United States. To our best knowledge, this is the first app implementing BAA. BAA first uses machine learning to construct a predictive quantitative model for each participant based on the historical step and goal data, and then, it uses the estimated model to generate challenging yet realistic step goals in an adaptive fashion by choosing step goals that, based on the estimated model, would maximize future physical activity. The primary aim of this RCT was to evaluate the efficacy of the automated mobile phone–based personalized,

adaptive goal-setting intervention as compared with the active control with nonpersonalized, steady daily step goals of 10,000. The main outcome measure was the relative change in objectively measured daily steps between the run-in period and 10 weeks. Secondary outcome measures included the following: step goal attainment (ie, fraction of step goals achieved by each participant), weight and height, self-reported sociodemographic information, self-reported medical history, Barriers to Being Active Quiz [48], and the short version of the international physical activity questionnaire [49]. We collected these survey results to investigate if the goal-setting component alone is capable of changing participants' survey responses before and after the study.

Methods

Study Design

The Cal Fitness study was a 10-week RCT with 2 groups: (1) the intervention group received automated personalized daily step goals, and (2) the control group received fixed daily step goals of 10,000 steps per day. The study was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley (UCB; institutional review board number 2016-03-8609), in July 2016 and was registered with the clinicaltrials.gov (NCT02886871) in August 2016. All participants provided written informed consent before study enrollment. This RCT was conducted in 2016 and analyzed in 2017.

Participant Recruitment

A total of 64 adult staff employees of UCB were recruited via email announcements. Recruitment commenced in August 2016 and ended in September 2016. The study ended in December 2016 to allow a 10-week period to all participants. Potential participants were contacted through email and then directed to a Web-based screening survey to assess eligibility. Those participants who met all the inclusion criteria were then contacted by trained study personnel via email to arrange an in-person session. Ineligible participants were informed by email to advise them that they are ineligible, and corresponding data were deleted.

The inclusion and exclusion criteria for the Cal Fitness study are given in Textbox 1.

Study Procedure

Eligible participants were asked to attend two 15-min in-person sessions (initial and 10-week post intervention visits) at UCB. The first in-person session occurred in September 2016, and the second session occurred in December 2016. During the first in-person session, a trained research staff member installed the CalFit app on participants' phones and advised the participants to keep the phone in their pocket or purse for the following 10-week period. A trained research staff member measured height (cm) and weight (kg) in both the sessions using a Seca 700 Physician's Balance Beam Scale with Height Rod, and body mass index (BMI) was also calculated. Participants were then instructed to complete the sociodemographic survey, the medical history survey, the Barriers to Being Active Quiz [37], and the short version of the international physical activity questionnaire [38]. During the second in-person session, a trained research staff removed the CalFit app from participants' phones. Participants were then instructed to complete the Barriers to Being Active Quiz [37] and the short version of the international physical activity questionnaire [38]. Participants received a US $50 Amazon gift certificate at completion if they completed all study requirements.

Textbox 1. Inclusion and exclusion criteria for the Cal Fitness study.

Inclusion criteria

- Staff member of University of California, Berkeley
- Intent to become physically active in the next 10 weeks, which was evaluated by asking potential participants if they wanted to increase their physical activity beyond their self-assessed current level
- Own an iPhone 5s (or a newer model)
- Willing to keep the iPhone in pockets during the day
- Willing to install and use the study app (which requires Internet connection) every day for 10 weeks
- Ability to speak and read English

Exclusion criteria

- Known medical conditions or physical problems that require special attention in an exercise program
- Planning an international trip during the next 3 months, which could interfere with daily server uploads of mobile phone data
- Pregnant or gave birth during the past 6 months
- Severe hearing or speech problem
- History of an eating disorder
- Current substance abuse
- Current participation in lifestyle modification programs or research studies that may confound study results
- History of bariatric surgery or plans for bariatric surgery in the next 12 months

CalFit iOS App

Our research team developed the CalFit app (iOS version, Apple Inc), which was designed to increase physical activity by allowing participants to track their daily step goals and to compare their step counts with their daily step goals in the past. Figure 1 shows the interface of the CalFit app. After participants open the app, they see the landing page and then the home page. On the home page, *number of steps completed today* and *today's step goal* are shown. Participants can click on two icons at the bottom of the home page. If they click on the left icon, the history page is displayed. The history page allows participants to track their performance over the past week by showing their daily steps and daily goals on a color-coded bar graph. The green bar indicates the accomplishment of achieving the step goal on the corresponding day, and the red bar indicates failure to achieve the step goal on the corresponding day. When participants click on the right icon, they reach the contact page that allows them to send messages to the research team. The built-in health chip in the iPhone collects the step data, and the accuracy of step counts collected by the iPhone health chip has been validated in a number of studies to have comparable accuracy to an ActiGraph [50-55]. One of these studies [51] conducted a large number of experiments and concluded that iPhones are accurate for tracking step counts, with a relative difference in the mean step count of −6.7% to 6.2% compared with direct observation. Another study [55] compared iPhone pedometer measurements with measurements from wearable devices in a free-living setting and concluded "measurements of number of steps and distance were excellent and could provide reliable judgment on the individuals' activity amount." Our app first saves the step and goal data locally on the phone and then syncs with the server every 10 min when the phone is active. The push notification for the app is also activated, and the standard iOS push notification is used. The push notification is visible in the landing page and in the recent notifications tab on the phone.

Run-In Period and Randomization

A total of 64 eligible participants started a 1-week run-in period after completing the initial in-person session. The purpose of the run-in period was to collect run-in daily steps, and assess if the participant was able to comply with the requirements needed to regularly use the CalFit app. During the run-in period, all participants in the control and the intervention groups received the identical set of daily step goals for day 1 to day 7 as 3000, 3500, 4000, 4500, 5000, 5500, and 6000 steps, respectively. The BAA algorithm was not used to compute step goals for participants in the intervention group during the run-in period. Dynamically increasing step goals were used in the run-in period to engage participants in using the app regularly. In addition, all participants received a push notification at 8 AM that provided today's step goal, and if the participant accomplished the goal before 8 PM, then another push notification was sent to congratulate the participant on reaching their step goal for the day. The identical goals between the 2 groups during the run-in period is used to establish a reference level of initial physical activity, which we used in our statistical analyses to compare the difference in daily steps between run-in and 10 weeks for the 2 groups. Data collected during the run-in period were used by the BAA algorithm to compute step goals for the intervention period. This is a valid approach because run-in data were indicative of the preference of different participants. All 64 participants were randomized to one of the 2 groups with a one-to-one ratio by a computer-based random number generator using the simple randomization approach. A one-to-one ratio means that each participant had a 50% probability of being assigned to one of the 2 groups, and the number of participants in each group may differ due to chance. The randomization to groups was implemented by the CalFit app after the run-in period, and the participants were aware of the 2 groups.

Figure 1. The CalFit app interface. (a) The landing page; (b) The homepage showing the steps done today and today's goal; (c) The "History" tab showing the performance of the past week. The black bar is the goal, and the bars are green for achieved goals and red for unachieved goals; (d) The "Contact Us" tab where participants can easily send messages to the study team.

Control

After the 1-week run-in period, participants in the control group were provided with constant daily step goals that were set to 10,000 steps per day through the CalFit app. Participants received a push notification at 8 AM every day that provided that day's step goal (10,000 steps), and if the participant achieved the goal before 8 PM, then another push notification was sent to congratulate the participant on reaching their step goal (of 10,000 steps) for the day.

Intervention

After the 1-week run-in period, participants in the intervention group received adaptively personalized step goals through the CalFit app. The daily step goals were computed using the BAA [43,47] on the complete history (past steps and goals) of the user. The BAA algorithm was applied every week to reduce variance in future steps and goals. Participants received a push notification at 8 AM every day that provided today's step goal, and if the participant accomplished the goal before 8 PM, then another push notification was sent to congratulate the participant on reaching their step goal for the day.

A rigorous mathematical formulation of the BAA algorithm that we used is provided in 2 studies [43,47]. This algorithm uses statistics and machine learning to adaptively compute personalized step goals that are predicted to maximize future physical activity for each participant based on all the past steps' data and goals of each participant. The BAA algorithm is applied to each participant individually, and it consists of two main steps. The first step is to use all of the participant's data to construct a quantitative model that predicts how many steps the participant will take in the future, given a prescribed set of step goals, and an important aspect of the model is a component that describes how achieving goals in the present can increase the likelihood of achieving goals in the future. The second step is to use this quantitative model to select a sequence of step goals that maximizes the predicted future number of steps. To make the process of updating step goals adaptive, the BAA algorithm is applied each week (using all the users' past data) to generate step goals for the coming week; moreover, the step goals computed by the BAA algorithm for the coming week are not constant, but increase or decrease based on the model prediction. A computer simulation study [43] found that applying the algorithm weekly is as effective as applying the algorithm daily (because steps can vary significantly on a day-to-day basis), and so we applied the algorithm weekly. More details about the BAA algorithm are provided in 2 studies [43,47].

Outcome Measures

The primary outcome of the study was the relative change in daily steps from run-in to the 10-week follow-up, measured objectively by the participants' iPhones. The daily step values were compared in the manner described in the statistical analysis section. Step count data were stored in a database on a private computer server at UCB. Data were automatically synced with the iPhone once every 10 min during the study. At the 10-week in-person session, complete step data were downloaded from the iPhone to store step count data that were unable to be transmitted. Data were unable to be transmitted if the app was turned off or no Internet connection was available. Other measures included weight and height, self-reported sociodemographic information, self-reported medical history, Barriers to Being Active Quiz [48] (which consists of 21 questions on a 10-point Likert scale on 7 subareas: lack of time, social influence, lack of energy, lack of willpower, fear of injury, lack of skill, and lack of resources), and the short version of the international physical activity questionnaire [49].

Statistical Analysis

Assuming an expected loss to follow-up of 10%, a target sample size of 30 participants per group was selected to give 80% power to detect between-group difference of 1500 steps with a pooled standard deviation (SD) of 2000 using a two-sided test and an alpha of .05. Differences between groups in run-in and 10 weeks were assessed using Student *t* test. The statistical analysis of the primary outcome of daily steps was performed using a linear mixed-effects model (LMM) with piecewise linear growth curve [56-58] with random effects for each individual of random slope and random intercept, and fixed effects of time, treatment group, and interaction term of time and treatment group. Our statistical analysis of the secondary outcome of step goal attainment (ie, fraction of step goals achieved by each participant) was performed by a similar LMM but with an additional specification of a binary response variable (ie, goal is either attained or not attained by an individual on a particular day). Means with 95% confidence intervals were obtained from the LMM. Sensitivity analysis was performed to obtain adjusted estimates of the effect of the treatment with the missing data on primary outcome, evaluated at $P<.05$. The primary cause of missing step data was failure to turn on the app. LMM implicitly imputes missing data by interpolation and is a common approach to deal with missing data in physical activity interventions [56-60]. (We did not use the common imputation method of "last observation carried forward" because it would increase bias in this context and lead to potentially false conclusions by inflating step counts at 10 weeks.) For accurate comparison between the control and the intervention groups, the weekly average steps in run-in were adjusted by adding the coefficient corresponding to each group (ie, control or intervention) computed by the LMM model. In addition, weekly moving average steps were computed by taking the average of each moving window with length 7, to reduce noise for better visualization.

To quantify app use for the intervention, a participant was categorized as a nonfrequent app user if the app was not used for a consecutive period of 7 days. By this criterion, 17 participants out of 34 in the intervention group and 16 participants out of 30 in the control group were frequent app users. Per-protocol analysis was performed on the 33 frequent app users, and intention-to-treat analysis was performed on all 64 subjects. Although the power for the per-protocol analysis will be low, the reason for conducting this analysis is that we want to investigate the impact of the CalFit app on an active subgroup, which could be more representative for its true performance if adopted in other full interventions that include additional components of a behavior change intervention. Intention-to-treat analysis was performed for the primary and secondary outcomes, and per-protocol analysis was performed only for the primary outcome. Missing survey response data

resulting from lost to follow-up was imputed by the latest available survey response of the subject. The statistical analysis was performed in MATLAB (MathWorks, Massachusetts, USA) version 9.0 [61] and R (R Core Team, Vienna, Austria) version 1.0.136 [62] in the year 2017.

Results

Recruitment Results

As shown in Figure 2, 97 potential participants were screened for eligibility by an online form, and 64 completed the initial in-person session.

Baseline Characteristics

Table 1 shows the baseline characteristics of the participants. A total of 34 participants were randomly assigned to the intervention group, and 30 participants were randomly assigned to the control group. All participants were included in the analysis based on the original assigned groups. Overall mean age was 41.1 (SD 11.3) years, and 83% (53/64) participants were female. In addition, 55% (35/64) of the participants self-identified as a member of a racial minority group. The baseline mean weight of participants was 77.2 kg (SD 18.7 kg) and the mean BMI was 27.3 kg/m^2(SD 6.1 kg/m^2). The mean height and weight for male and female participants were 177.5 cm and 82.6 kg and 165.9 cm and 76.1 kg, respectively. Furthermore, 20% of the participants reported at least one medical condition (ie, high blood pressure, type 2 diabetes, type 1 diabetes, coronary heart disease, or hypercholesterolemia). No baseline characteristics differed between the control and intervention groups. The run-in mean daily steps in the control and intervention groups were similar (7427 steps vs 7237 steps, respectively; *P*=.79) and are in line with baseline steps in other similar studies [14,42,63,64]. As shown in Multimedia Appendix 1, the self-reported survey results did not differ considerably between the 2 groups except for the lack of resources, which is a subscale of the Barriers to Being Active measure. The intervention group had a significantly higher rating of lack of resources than the control group (*P*=.03). We suspect this significant difference for lack of resources occurred due to chance.

Figure 2. Screening, randomization, and assessments of study participants.

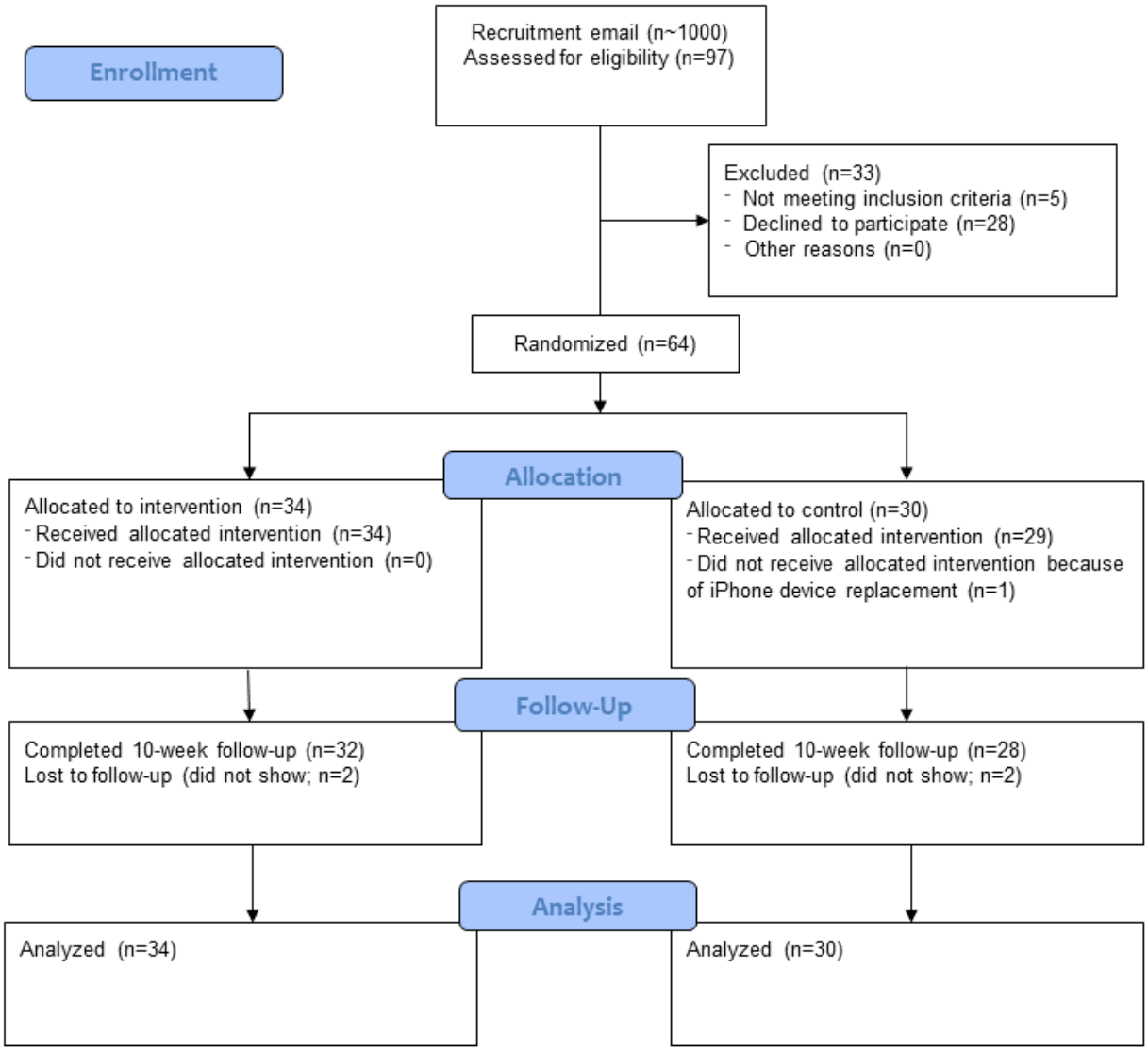


Table 1. Baseline characteristics between the control and intervention groups.

Baseline characteristics	All participants (N=64)	Control (N=30)	Intervention (N=34)	*P*
Run-in daily average steps, mean (SD)	7326 (2907)	7427 (2398)	7237 (3326)	.79
Age, years, mean (SD)	41.1 (11.3)	40.5 (10.5)	41.6 (12.2)	.72
Weight, kg, mean (SD)	77.2 (18.7)	77.8 (21.3)	77.0 (17.1)	.87
BMI[a], kg/m^2, mean (SD)	27.3 (6.1)	27.1 (6.7)	27.4 (5.8)	.82
Gender, n (%)				.82
Male	11 (17)	6 (20)	5 (15)	
Female	53 (83)	24 (80)	29 (85)	
Ethnicity, n (%)				.86
Asian	13 (20)	7 (23)	6 (18)	
Black or African American	8 (13)	3 (10)	5 (15)	
Hispanic or Latino	9 (14)	5 (17)	4 (12)	
White or non-Hispanic	29 (45)	13 (43)	16 (47)	
Other	5 (8)	2 (7)	3 (9)	
Marital status, n (%)				.20
Currently married or cohabitating	36 (56)	15 (50)	21 (62)	
Never married	21 (33)	13 (43)	8 (24)	
Divorced or widowed	7 (11)	2 (7)	5 (15)	
Education, n (%)				.30
Completed some college	5 (8)	1 (3)	4 (12)	
Completed college (4 years)	28 (44)	12 (40)	16 (47)	
Completed graduate school	31 (48)	17 (57)	14 (41)	
Work hour (per week), n (%)				.17
1-20 hours	3 (5)	3 (10)	0 (0)	
21-40 hours	16 (25)	7 (23)	9 (27)	
>40 hours	45 (70)	20 (67)	25 (74)	
Own a dog, n (%)				.99
Yes	16 (25)	8 (27)	8 (24)	
No	48 (75)	22 (73)	26 (77)	
Transportation to work, n (%)				.49
Car	28 (44)	10 (33)	18 (53)	
Public transportation	25 (39)	14 (47)	11 (32)	
Walk	4 (6)	2 (7)	2 (6)	
Bicycle	6 (9)	3 (10)	3 (9)	
Other	1 (2)	1 (3)	0 (0)	
Gym membership, n (%)				.45
Yes	32 (50)	13 (43)	19 (56)	
No	32 (50)	17 (57)	15 (44)	
Self-reported medical history, n (%)				
High blood pressure				.88
Yes	5 (8)	3 (10)	2 (6)	
No	59 (92)	27 (90)	32 (94)	
Type 2 diabetes				.43

Baseline characteristics	All participants (N=64)	Control (N=30)	Intervention (N=34)	*P*
Yes	5 (8)	1 (3)	4 (12)	
No	59 (92)	29 (97)	30 (88)	
Type 1 diabetes				.62
Yes	0 (0)	0 (0)	0 (0)	
No	64 (100)	30 (100)	34 (100)	
Coronary heart disease				.62
Yes	0 (0)	0 (0)	0 (0)	
No	64 (100)	30 (100)	34 (100)	
Hypercholesterolemia				.83
Yes	7 (11)	4 (13)	3 (9)	
No	53 (83)	24 (80)	29 (85)	
Unknown	4 (6)	2 (7)	2 (6)	

[a]BMI: body mass index

Efficacy of Intervention

Main Analysis

Intention-to-treat analyses indicated that the intervention group had a decrease in mean (SD) daily step count of 390 (SD 490) steps between run-in and 10 weeks compared with a decrease of 1350 (SD 420) steps among controls (*P*=.03). The net difference in daily steps between the groups was 960 steps (95% CI 90-1830 steps). Table 2 shows the run-in adjusted objectively measured raw average weekly steps for both the groups without missing data imputation. Figure 3 shows the run-in adjusted weekly average steps and moving average steps for intention-to-treat.

The average step goals for the first week are the same for both the control and the intervention groups because both received the same goals during the first week. Table 3 gives the fraction of achieved step goals for the 2 groups. Intention-to-treat analysis indicated that the intervention group had a decrease in mean fraction of achieved step goals of 0.34 (SD 0.05) between run-in and 10 weeks compared with a decrease of 0.49 (SD 0.04) among controls (*P*=.003). The net difference in fraction of achieved step goals between the groups was 0.15 (95% CI 0.02-0.25). Figure 4 details the intention-to-treat weekly average step goals and the fraction of achieved step goals for the 2 groups.

Sensitivity Analysis

Per-protocol analysis (among the 33 frequent app users: 16 in control and 17 in intervention groups) indicated that the intervention group had a decrease in mean (SD) daily step count of 0 (SD 420) steps between run-in and 10 weeks, whereas the control group had a decrease of 1500 (SD 550) steps (*P*=.03). The net difference in daily steps between the groups was 1500 steps (95% CI 130-2900 steps). Figure 5 shows the run-in adjusted weekly average steps and moving average steps for per-protocol.

Per-protocol analysis also indicated that the intervention group had a decrease in mean (SD) fraction of achieved step goals of 0.27 (SD 0.08) between run-in and 10 weeks compared with a decrease of 0.46 (SD 0.06) among controls (*P*=.02). The net difference in fraction of achieved step goals between the groups was 0.19 (95% CI 0.02-0.38). Figure 6 details the per-protocol weekly average step goal and the fraction of achieved step goals for the 2 groups.

Other Analysis

No significant difference (Multimedia Appendix 1) in self-reported physical activity scores and Barriers to Being Active was noted within the 2 groups over time, and no significant difference between the 2 groups was observed at run-in or 10 weeks.

Fidelity of In-Person Sessions

The number of participants in the 2 groups who failed to complete the second in-person session for follow-up did not differ (*P*=.90): 6.6% (n=2) in the control group and 5.9% (n=2) in the intervention group (Figure 2). Their data were included in the analysis.

Table 2. Run-in adjusted objectively recorded (using iPhone) physical activity.

Week	Mean number of steps	
	Control (N=30)	Intervention (N=34)
Run-in	7462	7623
Week 2	7674	7882
Week 3	7650	7290
Week 4	7834	8094
Week 5	7494	7611
Week 6	7183	6958
Week 7	7308	7399
Week 8	6770	7237
Week 9	6855	7129
Week 10	6471	7549

Figure 3. Weekly average and moving average steps for the 2 groups over the course of the study for intention-to-treat analysis after run-in adjustment. Left panel: mean weekly steps for intention-to-treat; Right panel: weekly moving average for intention-to-treat.

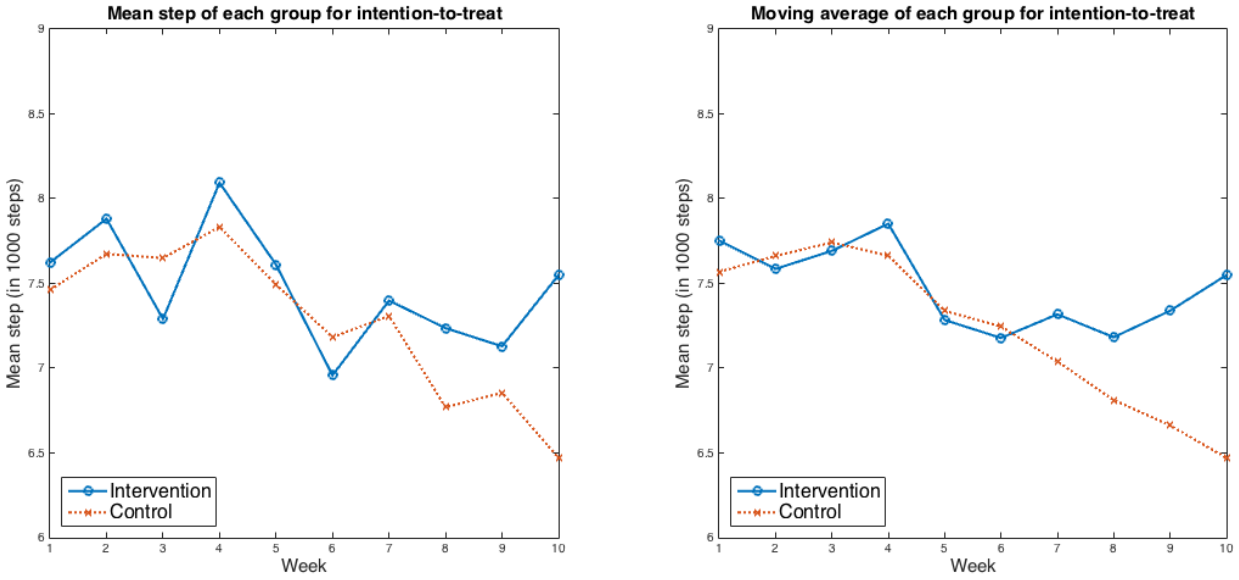


Table 3. Fraction of achieved daily step goals in weeks.

Week	Control (N=30)	Intervention (N=34)
Week 1 (run-in)	0.74	0.71
Week 2	0.34	0.49
Week 3	0.34	0.41
Week 4	0.29	0.44
Week 5	0.28	0.34
Week 6	0.25	0.33
Week 7	0.29	0.37
Week 8	0.23	0.34
Week 9	0.21	0.36
Week 10	0.19	0.34

Figure 4. Weekly average step goals and average fraction of goals achieved for the 2 groups for intention-to-treat analysis. Left panel: weekly average step goals for intention-to-treat; Right panel: weekly average fraction of achieved goals for intention-to-treat.

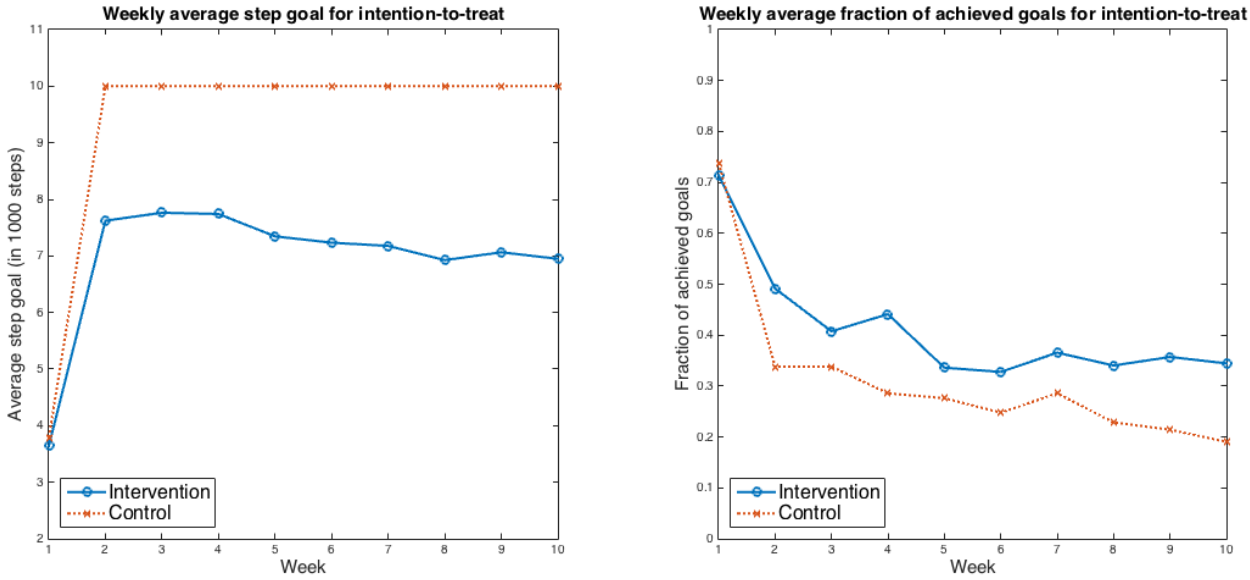


Figure 5. Weekly average and moving average steps for the 2 groups over the course of the study for per-protocol analysis after run-in adjustment. Left panel: mean weekly steps for per-protocol; Right panel: weekly moving average for per-protocol.

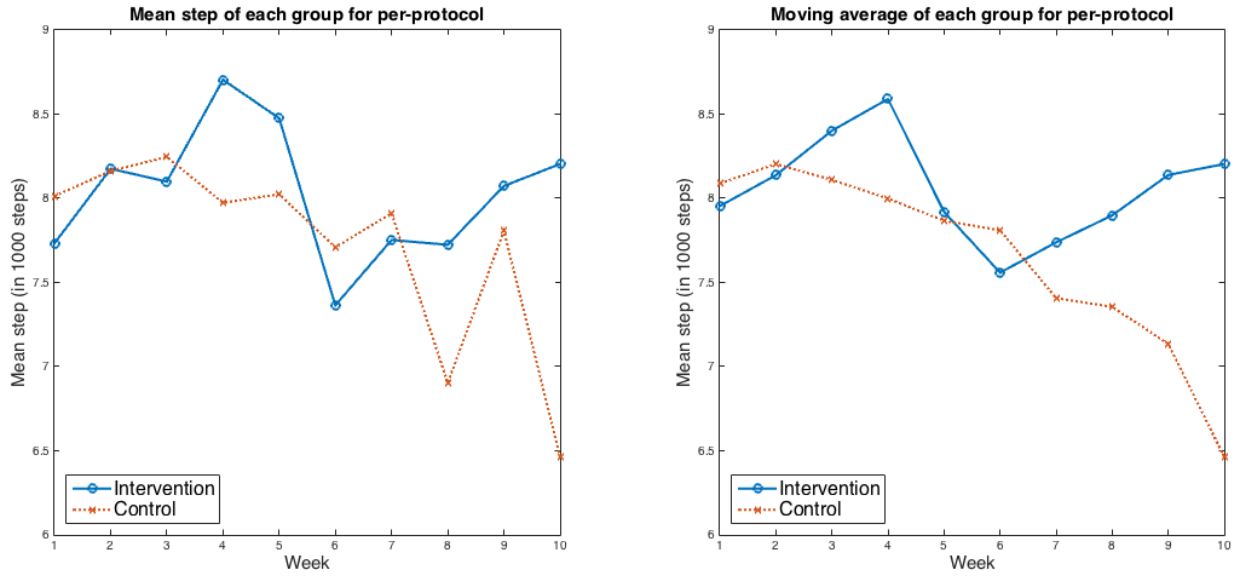
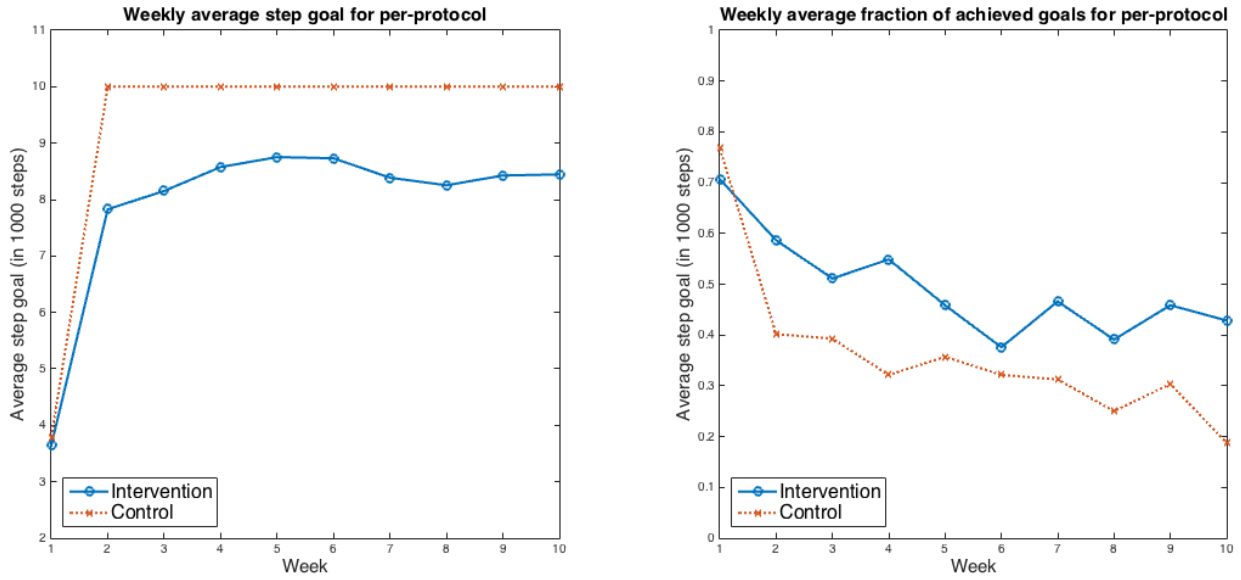


Figure 6. Weekly average step goals and average fraction of goals achieved for the 2 groups for per-protocol analysis. Left panel: weekly average step goals for per-protocol; Right panel: weekly average fraction of achieved goals for per-protocol.



Discussion

Principal Findings

This study evaluated the efficacy of a mobile phone–based physical activity intervention that provided adaptively personalized daily step goals. The intervention led to a statistically significant difference of 960 more daily steps in the intervention group compared with the control group over 10 weeks, in line with similar studies [28,29]. Although both groups had reduced daily steps at 10 weeks as compared with run-in, we speculate this was caused by run-in step counts being higher than the natural baseline. We believe this inverse relationship was a result of participants receiving step goals and monitoring step count through the CalFit app or the built-in iPhone Health app during the run-in period. This is supported by the observations that during the run-in period, all participants received daily step goals of 3000, 3500, 4000, 4500, 5000, 5500, and 6000 steps and initially over responded to these goals, and that the trends in daily steps between the control and intervention groups began to diverge in the 6th week of the study when enthusiasm of study participants wore out. Thus, later in the study, the personalized daily step goals seemed to be more effective in engaging participants and maintaining daily step counts compared with constant step goals.

The health literature has identified that setting goals is effective in lifestyle modification and physical activity promotion [9,14,36,65]. One analysis found that the importance of goal attainment and self-efficacy are the two main factors that contribute to goal commitment [32]. More recent studies [66-68] showed that individuals with higher self-efficacy are more likely to achieve activity goals and that failing to achieve activity goals reduces individuals' self-efficacy. Therefore, activity goals need to be set with care. Past studies [69-71] and most persuasive technologies [34,72] either adopted a steady goal of 10,000 steps or allowed self-set goals. To our knowledge, this is the first study to use machine learning to automatically set adaptively personalized step goals and deliver the step goals using a mobile phone technology. The RCT outcomes show that adaptively personalized goals were important in promoting physical activity relative to constant step goals. The adaptive step goals were set to be challenging yet attainable; thus, the average step goals for the intervention group were lower than the average step goals for the control group. As the adaptive step goals were designed to be challenging, the goal achieving percentage for the intervention group was not 100%. Instead, we observed the goal achieving percentage for the intervention group was 30%-40%, which was 15% more compared with the goal achieving percentage for the control group. Being able to achieve more daily step goals can enhance participants' self-efficacy, which further promotes physical activity in the days to follow [32,73-75]. The significantly higher (but not too high) rate of achieving step goals and significantly more steps of the participants in the intervention group demonstrate that the BAA algorithm computed adaptively personalized step goals that were capable of being both challenging and manageable for participants, and these goals effectively promoted physical activity.

Nonadherence is another challenge in mobile phone–based lifestyle modification programs. As a result, many past mHealth interventions involve regular in-person counseling sessions besides the mobile intervention to motivate adherence [9,14,76]. However, in-person counseling sessions are costly and put a burden on both the participants and the research staff [77-79]. Our study was intentionally designed to have only two in-person sessions (each of 15 min) at run-in and at 10 weeks to better simulate the environment of a completely mobile phone–based physical activity intervention. Note that the two in-person sessions in this study were necessary in-person contacts for the purpose of assessment in the study; they are different from in-person counseling sessions that serve as an essential part of an actual intervention. Despite the absence of coaching sessions, the percentage of frequent users observed over 10 weeks in our study was better than that reported in similar trials [71,80,81].

Our results indicate that a mobile phone–based intervention without coaching sessions is still effective in promoting physical activity. In-person contact and coaching sessions are therefore not necessary requirements for effective physical activity interventions, and there is potential to replace those contacts with better-designed physical activity apps.

An additional advantage of this study is that it only relied on one device for both data collection and intervention delivery. Similar studies either used a pedometer or accelerometer besides the mobile phone or requested regular data inputs from the participants, requiring greater efforts on the participant side, which was shown to be burdensome and could lead to declining use of the app [82]. In this study, step data were objectively measured by the iPhone, and participants were only requested to carry their mobile phones with them (in their pocket or their purse). No other manual data entry was needed on the participant side. Moreover, the CalFit app is designed in a flexible way that is compatible with other data collection devices, such as wearable step trackers, as long as the step data can be synced with the iPhone.

In addition to objectively measured outcomes, it is of interest to investigate if self-reported survey results differ between this study and full behavioral interventions (with many behavior change components). Barriers to Being Active quiz and the international physical activity questionnaire are popular surveys that have been widely adopted [14,33,81,83-85]. Researchers found that there exist significant differences in survey responses before and after a full behavioral intervention [14,33,86,87]. However, we failed to observe such difference. We suspect that goal setting alone may not be strong enough to change participants' opinion on self-reported surveys and that other behavior change components are required (eg, coaching sessions).

This study tested one single component of behavior change (ie, goal setting), and the purpose of this design was to isolate the impact of goal setting from other behavior change components. Beyond goal setting, there are many other components of behavior change that can be beneficial for fitness apps. For instance, customized messages and social interactions have the potential to further improve the efficacy of fitness apps. This study is not designed to be a stand-alone intervention but rather to provide evidence on the efficacy of evaluating a single design component to motivate future evaluations on other design components. We believe there is great potential for better-designed fitness apps that can contribute to more effective physical activity intervention.

Limitations

The first limitation of this study is the relatively small sample size, which only contained UCB adult staff workers with a dominant proportion of females (83% [53/64]). The results may not generalize to the general public. The relatively high education level of the participants may also limit the generalizability. In addition, the CalFit app was only available on the iOS platform, which could bias results. Second, the daily steps assessment during the run-in period was not able to establish a natural baseline. Therefore, our trial could only determine the relative (to the control) benefit of the intervention, but could not determine the absolute (compared with the natural baseline) benefit of the intervention. Blocked display of step counts with no step goal during the run-in period may provide additional insights to the natural baseline. Third, the iPhone was not able to collect data when it was being turned off or was not with the participant, and it was not able to distinguish the carrying method (purse vs pocket). However, the chance of the above happening was the same for the control and the intervention groups because of randomization; so, these factors do not impact the relative step differences between the 2 groups. Fourth, this study did not assess the underlying behavior skills (self-efficacy, goal setting, etc) that may impact individual's response to interventions. Finally, the study was conducted for 10 weeks, which is a relatively short time. Studies that span a longer period are needed to evaluate the long-term effect of such personalized step goal-setting intervention delivered via mobile phones.

Conclusions

Our RCT indicates that mobile phone–delivered adaptively personalized step goals are promising in promoting physical activity. The intervention led to a statistically significant difference of 960 more daily steps in the intervention group compared with the control group over 10 weeks. The higher (but not too high) percentage of goal achievement in the intervention group confirms that the adaptively personalized step goals computed by the BAA algorithm used in this trial are capable of creating challenging yet attainable goals. The significant step difference between the 2 groups suggests that a mobile phone–based physical activity intervention with reduced in-person sessions is feasible. The results obtained in this study can guide the design of future mobile phone–based physical activity interventions.

Acknowledgments

The authors would like to thank Emily Ma, Smita Jain, and Jessica Lin for their help during the prestudy in-person session. The authors would also like to thank Mingyang Li and Alejandro Castillejo for their help in developing the app. AA and YM were supported in part by funding from the Philippine-California Advanced Research Institutes (PCARI). MZ and KG were supported in part by funding from the UC Center for Information Technology Research in the Interest of Society and the PCARI grant IIID-2015-07. YF's effort for this project was in part supported by a grant (K24NR015812) from the National Institute of Nursing Research and a grant (R01HL104147) from the National Heart, Lung, and Blood Institute. EF's effort for this project was supported by a grant from the National Center for Advancing Translational Sciences of the National Institutes of Health (KL2TR000143). The study sponsors had no role in the study design; collection, analysis, or interpretation of data; writing the report; or the decision to submit the report for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Self-reported physical activity scores and barriers to being active of the two groups pre and post the study.

[PDF File (Adobe PDF File), 41KB - mhealth_v6i1e28_app1.pdf]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 734KB - mhealth_v6i1e28_app2.pdf]

References

1. World Health Organization. Health topics: physical activity URL: http://www.who.int/topics/physical_activity/en/ [WebCite Cache ID 6tnhpRHzS]
2. Knight J. Physical inactivity: associated diseases and disorders. Ann Clin Lab Sci 2012;42(3):320-337. [Medline: 22964623]
3. Thompson PD, Buchner D, Pina IL, Balady GJ, Williams MA, Marcus BH, et al. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease. Circulation 2003 Jun 24;107(24):3109-3116 [FREE Full text] [doi: 10.1161/01.CIR.0000075572.40158.77] [Medline: 12821592]
4. Sigal RJ, Kenny GP, Wasserman DH, Castaneda-Sceppa C, White RD. Physical activity/exercise and type 2 diabetes: a consensus statement from the American Diabetes Association. Diabetes Care 2006 Jun;29(6):1433-1438. [doi: 10.2337/dc06-9910] [Medline: 16732040]
5. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm 2009 Jun;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
6. Health.gov. 2008 Physical activity guidelines for Americans URL: https://health.gov/paguidelines/pdf/paguide.pdf [WebCite Cache ID 6tniVY36p]
7. Heart.org. Physical inactivity statistical fact sheet URL: https://www.heart.org/idc/groups/heart-public/@wcm/@sop/@smd/documents/downloadable/ucm_319589.pdf [WebCite Cache ID 6tniaH2X2]
8. State of Obesity. Physical inactivity in the United States URL: https://stateofobesity.org/physical-inactivity/ [WebCite Cache ID 6tniduS1K]
9. Eakin EG, Glasgow RE, Riley KM. Review of primary care-based physical activity intervention studies: effectiveness and implications for practice and future research. J Fam Pract 2000 Feb;49(2):158-168. [Medline: 10718694]
10. Little P, Dorward M, Gralton S, Hammerton L, Pillinger J, White P, et al. A randomised controlled trial of three pragmatic approaches to initiate increased physical activity in sedentary patients with risk factors for cardiovascular disease. Br J Gen Pract 2004 Mar;54(500):189-195 [FREE Full text] [Medline: 15006124]
11. Pahor M, Blair SN, Espeland M, Fielding R, Gill TM, Guralnik JM, et al. Effects of a physical activity intervention on measures of physical performance: results of the lifestyle interventions and independence for elders pilot (LIFE-P) study. J Gerontol A Biol Sci Med Sci 2006 Nov;61(11):1157-1165. [Medline: 17167156]
12. Fritz T, Huang E, Murphy G, Zimmermann T. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; 2014; Toronto p. 487-496.
13. Fjeldsoe B, Miller Y, Marshall A. MobileMums: a randomized controlled trial of an SMS-based physical activity intervention. Ann Behav Med 2010 May;39(2):101-111. [doi: 10.1007/s12160-010-9170-z] [Medline: 20174902]
14. Fukuoka Y, Vittinghoff E, Jong SS, Haskell W. Innovation to motivation--pilot study of a mobile phone intervention to increase physical activity among sedentary women. Prev Med 2010;51(3-4):287-289 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.006] [Medline: 20600263]
15. Hurling R, Catt M, De Boni M, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr 27;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
16. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]
17. O'Reilly GA, Spruijt-Metz D. Current mHealth technologies for physical activity assessment and promotion. Am J Prev Med 2013 Oct;45(4):501-507 [FREE Full text] [doi: 10.1016/j.amepre.2013.05.012] [Medline: 24050427]
18. van den Berg MH, Schoones JW, Vliet Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007 Sep 30;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]

19. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. J Cardiovasc Nurs 2013;28(4):320-329 [FREE Full text] [doi: 10.1097/JCN.0b013e318250a3e7] [Medline: 22635061]
20. Fanning J, Mullen S, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012 Nov 21;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
21. Araiza P, Hewes H, Gashetewa C, Vella CA, Burge MR. Efficacy of a pedometer-based physical activity program on parameters of diabetes control in type 2 diabetes mellitus. Metabolism 2006 Oct;55(10):1382-1387. [doi: 10.1016/j.metabol.2006.06.009] [Medline: 16979410]
22. Hultquist CN, Albright C, Thompson DL. Comparison of walking recommendations in previously inactive women. Med Sci Sports Exerc 2005 Apr;37(4):676-683. [Medline: 15809569]
23. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. Am J Health Promot 2006;21(2):85-89. [Medline: 17152246]
24. Lindberg R. Active living: on the road with the 10,000 steps program. J Am Diet Assoc 2000 Aug;100(8):878-879. [Medline: 10955038]
25. Pila E, Tignor S, Gilchrist J, Sabiston C, Fombelle P, Sirianni N. Self-conscious emotions in response to physical activity success and failure: findings from a global 112-day pedometer intervention. Journal of Exercise, Movement, and Sport (SCAPPS refereed abstracts repository) 2016;48(1) [FREE Full text]
26. Ryan J, Edney S, Maher C. Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med 2017 Dec;7(4):702-708 [FREE Full text] [doi: 10.1007/s13142-017-0499-8] [Medline: 28523603]
27. Ransdell L, Robertson L, Ornes L, Moyer-Mileur L. Generations exercising together to improve fitness (GET FIT): a pilot study designed to increase physical activity and improve health-related fitness in three generations of women. Women Health 2004;40(3):77-94. [Medline: 15829447]
28. Adams M, Hurley J, Todd M, Bhuiyan N, Jarrett C, Tucker W, et al. Adaptive goal setting and financial incentives: a 2 × 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health 2017 Dec 29;17(1):286 [FREE Full text] [doi: 10.1186/s12889-017-4197-8] [Medline: 28356097]
29. Adams M, Sallis J, Norman G, Hovell M, Hekler E, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
30. Shilts M, Horowitz M, Townsend M. Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. Am J Health Promot 2004;19(2):81-93. [Medline: 15559708]
31. Locke E, Latham G. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [Medline: 12237980]
32. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
33. Fukuoka Y, Gay C, Joiner K, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]
34. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
35. Joseph R, Keller C, Adams M, Ainsworth B. Print versus a culturally-relevant facebook and text message delivered intervention to promote physical activity in African American women: a randomized pilot trial. BMC Womens Health 2015 Mar 27;15:30 [FREE Full text] [doi: 10.1186/s12905-015-0186-1] [Medline: 25886945]
36. Sidman CL, Corbin CB, Le MG. Promoting physical activity among sedentary women using pedometers. Res Q Exerc Sport 2004 Jun;75(2):122-129. [doi: 10.1080/02701367.2004.10609143] [Medline: 15209330]
37. Tudor-Locke C. Eric.ed. 2002. Taking steps toward increased physical activity: using pedometers to measure and motivate URL: https://files.eric.ed.gov/fulltext/ED470689.pdf [accessed 2018-01-16] [WebCite Cache ID 6wVQRniyz]
38. Williams B, Bezner J, Chesbro S, Leavitt R. The effect of a behavioral contract on adherence to a walking program in postmenopausal African American women. Top Geriatr Rehabil 2005;21(4):332-342.
39. Chan CB, Ryan DAJ, Tudor-Locke C. Health benefits of a pedometer-based physical activity intervention in sedentary workers. Prev Med 2004 Dec;39(6):1215-1222. [doi: 10.1016/j.ypmed.2004.04.053] [Medline: 15539058]
40. Poirier J, Bennett W, Jerome G, Shah N, Lazo M, Yeh HC, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Feb 09;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
41. Bravata D, Smith-Spangler C, Sundaram V, Gienger A, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
42. Croteau KA. A preliminary study on the impact of a pedometer-based intervention on daily steps. Am J Health Promot 2004;18(3):217-220. [Medline: 14748310]

43. Mintz Y, Aswani A, Kaminsky P, Flowers E, Fukuoka Y. Arxiv. 2017. Behavioral analytics for myopic agents URL: https://arxiv.org/abs/1702.05496 [accessed 2018-01-16] [WebCite Cache ID 6wVQdAyoZ]
44. Finkelstein EA, Brown DS, Brown DR, Buchner DM. A randomized study of financial incentives to increase physical activity among sedentary older adults. Prev Med 2008 Aug;47(2):182-187. [doi: 10.1016/j.ypmed.2008.05.002] [Medline: 18571226]
45. Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
46. Paul-Ebhohimhen V, Avenell A. Systematic review of the use of financial incentives in treatments for obesity and overweight. Obes Rev 2008 Jul;9(4):355-367. [doi: 10.1111/j.1467-789X.2007.00409.x] [Medline: 17956546]
47. Aswani A, Kaminsky P, Mintz Y, Flowers E, Fukuoka Y. SSRN. 2016. Behavioral modeling in weight loss interventions URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838443 [accessed 2018-01-16] [WebCite Cache ID 6wVQjRlK8]
48. CDC. Barriers to being active quiz URL: http://www.cdc.gov/diabetes/ndep/pdfs/8-road-to-health-barriers-quiz-508.pdf [WebCite Cache ID 6tnjryG4a]
49. Craig C, Marshall A, Sjöström M, Bauman A, Booth M, Ainsworth B, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
50. Bai Y, Welk GJ, Nam YH, Lee JA, Lee J, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016 Jan;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
51. Case M, Burwick H, Volpp K, Patel M. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
52. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Dec 20;547(7663):336-339. [doi: 10.1038/nature23018] [Medline: 28693034]
53. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
54. Lee J, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. Med Sci Sports Exerc 2014 Sep;46(9):1840-1848. [doi: 10.1249/MSS.0000000000000287] [Medline: 24777201]
55. Wen D, Zhang X, Liu X, Lei J. Evaluating the consistency of current mainstream wearable devices in health monitoring: a comparison under free-living conditions. J Med Internet Res 2017 Mar 07;19(3):e68 [FREE Full text] [doi: 10.2196/jmir.6874] [Medline: 28270382]
56. Naumova E, Must A, Laird N. Tutorial in biostatistics: evaluating the impact of 'critical periods' in longitudinal studies of growth using piecewise mixed effects models. Int J Epidemiol 2001 Dec;30(6):1332-1341. [Medline: 11821342]
57. Tabák AG, Jokela M, Akbaraly T, Brunner E, Kivimäki M, Witte D. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. Lancet 2009 Jun 27;373(9682):2215-2221 [FREE Full text] [doi: 10.1016/S0140-6736(09)60619-X] [Medline: 19515410]
58. Phillips SM, Bandini LG, Naumova EN, Cyr H, Colclough S, Dietz WH, et al. Energy-dense snack food intake in adolescence: longitudinal relationship to weight and fatness. Obes Res 2004 Mar;12(3):461-472 [FREE Full text] [doi: 10.1038/oby.2004.52] [Medline: 15044663]
59. Verbeke G. Linear mixed models for longitudinal data. New York: Springer; 1997.
60. Krueger C, Tian L. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biol Res Nurs 2004 Oct;6(2):151-157. [doi: 10.1177/1099800404267682] [Medline: 15388912]
61. Mathworks. Natick, MA: The Mathworks Inc; 2016. MATLAB and ttatistics toolbox release 2016a URL: https://in.mathworks.com/products/new_products/release2016a.html [accessed 2018-01-16] [WebCite Cache ID 6wVRRrVNV]
62. R-project. Vienna, Austria: R Foundation for Statistical Computing; 2017. R: A language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2018-01-16] [WebCite Cache ID 6wVRZHFvn]
63. Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps per day in US adults. Med Sci Sports Exerc 2009 Jul;41(7):1384-1391. [doi: 10.1249/MSS.0b013e318199885c] [Medline: 19516163]
64. Stovitz SD, VanWormer JJ, Center BA, Bremer KL. Pedometers as a means to increase ambulatory activity for patients seen at a family medicine clinic. J Am Board Fam Pract 2005;18(5):335-343 [FREE Full text] [Medline: 16148243]
65. Annesi JJ. Goal-setting protocol in adherence to exercise by Italian adults. Percept Mot Skills 2002 Apr;94(2):453-458. [doi: 10.2466/pms.2002.94.2.453] [Medline: 12027338]
66. Barz M, Lange D, Parschau L, Lonsdale C, Knoll N, Schwarzer R. Self-efficacy, planning, and preparatory behaviours as joint predictors of physical activity: a conditional process analysis. Psychol Health 2016;31(1):65-78. [doi: 10.1080/08870446.2015.1070157] [Medline: 26155825]
67. Olander E, Fletcher H, Williams S, Atkinson L, Turner A, French D. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 03;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]

68. Iwasaki Y, Honda S, Kaneko S, Kurishima K, Honda A, Kakinuma A, et al. Exercise self-efficacy as a mediator between goal-setting and physical activity: developing the workplace as a setting for promoting physical activity. Saf Health Work 2017 Mar;8(1):94-98 [FREE Full text] [doi: 10.1016/j.shaw.2016.08.004] [Medline: 28344847]
69. Clarke KK, Freeland-Graves J, Klohe-Lehman DM, Milani TJ, Nuss HJ, Laffrey S. Promotion of physical activity in low-income mothers using pedometers. J Am Diet Assoc 2007 Jun;107(6):962-967. [doi: 10.1016/j.jada.2007.03.010] [Medline: 17524717]
70. Merom D, Rissel C, Phongsavan P, Smith BJ, Van KC, Brown WJ, et al. Promoting walking with pedometers in the community: the step-by-step trial. Am J Prev Med 2007 Apr;32(4):290-297. [doi: 10.1016/j.amepre.2006.12.007] [Medline: 17303369]
71. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 01;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
72. Wang J, Cadmus-Bertram L, Natarajan L, White M, Madanat H, Nichols J, et al. Wearable sensor/device (fitbit one) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792 [FREE Full text] [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
73. Bandura A, Cervone D. Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. J Pers Soc Psychol 1983;45(5):1017-1028. [doi: 10.1037/0022-3514.45.5.1017]
74. Anderson ES, Wojcik JR, Winett RA, Williams DM. Social-cognitive determinants of physical activity: the influence of social support, self-efficacy, outcome expectations, and self-regulation among participants in a church-based health promotion study. Health Psychol 2006 Jul;25(4):510-520. [doi: 10.1037/0278-6133.25.4.510] [Medline: 16846326]
75. French DP, Olander EK, Chisholm A, Mc SJ. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. Ann Behav Med 2014 Oct;48(2):225-234. [doi: 10.1007/s12160-014-9593-z] [Medline: 24648017]
76. Venditti EM, Wylie-Rosett J, Delahanty LM, Mele L, Hoskin MA, Edelstein SL, Diabetes Prevention Program Research Group. Short and long-term lifestyle coaching approaches used to address diverse participant barriers to weight loss and physical activity adherence. Int J Behav Nutr Phys Act 2014 Feb 12;11:16 [FREE Full text] [doi: 10.1186/1479-5868-11-16] [Medline: 24521153]
77. Cummins CO, Evers KE, Johnson JL, Paiva A, Prochaska JO, Prochaska JM. Assessing stage of change and informed decision making for Internet participation in health promotion and disease management. Manag Care Interface 2004 Aug;17(8):27-32. [Medline: 15471108]
78. Brug J, Oenema A, Kroeze W, Raat H. The internet and nutrition education: challenges and opportunities. Eur J Clin Nutr 2005 Aug;59(Suppl 1):S130-S139. [doi: 10.1038/sj.ejcn.1602186] [Medline: 16052181]
79. Wylie-Rosett J, Swencionis C, Ginsberg M, Cimino C, Wassertheil-Smoller S, Caban A, et al. Computerized weight loss intervention optimizes staff time: the clinical and cost results of a controlled clinical trial conducted in a managed care setting. J Am Diet Assoc 2001 Oct;101(10):1155-1162. [Medline: 11678486]
80. Atienza AA, King AC, Oliveira BM, Ahn DK, Gardner CD. Using hand-held computer technologies to improve dietary intake. Am J Prev Med 2008 Jun;34(6):514-518. [doi: 10.1016/j.amepre.2008.01.034] [Medline: 18471588]
81. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
82. Ahtinen A, Mattila E, Vaatanen A, Hynninen L, Salminen J, Koskinen E. User experiences of mobile wellness applications in health promotion: user study of wellness diary, mobile coach and selfrelax. 2009 Presented at: Pervasive Computing Technologies for Healthcare; 2009; London. [doi: 10.4108/ICST.PERVASIVEHEALTH2009.6007]
83. Kulavic K, Hultquist CN, McLester JR. A comparison of motivational factors and barriers to physical activity among traditional versus nontraditional college students. J Am Coll Health 2013;61(2):60-66. [doi: 10.1080/07448481.2012.753890] [Medline: 23409855]
84. Sawchuk C, Russo J, Bogart A, Charles S, Goldberg J, Forquera R, et al. Barriers and facilitators to walking and physical activity among American Indian elders. Prev Chronic Dis 2011 May;8(3):A63 [FREE Full text] [Medline: 21477503]
85. Kim Y, Park I, Kang M. Convergent validity of the international physical activity questionnaire (IPAQ): meta-analysis. Public Health Nutr 2013 Mar;16(3):440-452. [doi: 10.1017/S1368980012002996] [Medline: 22874087]
86. Bombardier C, Fann J, Ludman E, Vannoy S, Dyer J, Barber J, et al. The relations of cognitive, behavioral, and physical activity variables to depression severity in traumatic brain injury: reanalysis of data from a randomized controlled trial. J Head Trauma Rehabil 2017;32(5):343-353. [doi: 10.1097/HTR.0000000000000288] [Medline: 28195952]
87. Sandroff B, Klaren R, Pilutti L, Dlugonski D, Benedict R, Motl R. Randomized controlled trial of physical activity, cognition, and walking in multiple sclerosis. J Neurol 2014 Feb;261(2):363-372. [doi: 10.1007/s00415-013-7204-8] [Medline: 24323244]

Abbreviations

BAA: behavioral analytics algorithm
BMI: body mass index
iOS: iPhone operating system
LMM: linear mixed-effects model
mHealth: mobile health
PCARI: Philippine-California Advanced Research Institutes
RCT: randomized controlled trial
UCB: University of California, Berkeley

Edited by G Eysenbach; submitted 04.10.17; peer-reviewed by B Fjeldsoe, A Rickman, N Bragazzi, C Brakenridge, J Dallinga, H Nero; comments to author 02.11.17; revised version received 29.11.17; accepted 16.12.17; published 25.01.18.

Please cite as:
Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, Aswani A
Evaluating Machine Learning–Based Automated Personalized Daily Step Goals Delivered Through a Mobile Phone App: Randomized Controlled Trial
JMIR Mhealth Uhealth 2018;6(1):e28
URL: http://mhealth.jmir.org/2018/1/e28/
doi:10.2196/mhealth.9117
PMID:29371177

Original Paper

Using Mobile Health Intervention to Improve Secondary Prevention of Coronary Heart Diseases in China: Mixed-Methods Feasibility Study

Shu Chen[1], MSc; Enying Gong[1], MSc; Dhruv S Kazi[2,3,4], MSc, MD; Ann B Gates[5], BPharm (Hons); Rong Bai[6,7], MD; Hua Fu[8], PhD; Weixia Peng[8], PhD; Ginny De La Cruz[9], MSc; Lei Chen[10], MD; Xianxia Liu[10], MD; Qingjie Su[11], MD; Nicolas Girerd[12,13], MPH, MD, PhD; Kamilu M Karaye[14], FWACP, MBBS, PhD; Khalid F Alhabib[15], MBBS, FESC, FACC, FRCPC; Lijing L Yan[1,16], MPH, PhD; JD Schwalm[17,18], MD

[1]Global Health Research Center, Duke Kunshan University, Kunshan, China

[2]Department of Medicine (Cardiology), University of California, San Francisco, San Francisco, CA, United States

[3]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, United States

[4]Division of Cardiology, Zuckerberg San Francisco General Hospital, San Francisco, CA, United States

[5]Exercise Works Ltd., Derby, United Kingdom

[6]National Clinical Research Center for Cardiovascular Diseases, Beijing, China

[7]Department of Cardiology, Beijing Anzhen Hospital, Capital Medical University, Beijing, China

[8]School of Public Health, Fudan University, Shanghai, China

[9]Global Health Research Center, Duke Kunshan University, Kunshan, China

[10]Department of Cardiology, Hainan Provincial Nongken General Hospital, Haikou, China

[11]Department of Neurology, Hainan Provincial Nongken General Hospital, Haikou, China

[12]Institut Lorrain du Coeur et des Vaisseaux Louis Mathieu, Centre d'Investigations Cliniques, Centre Hospitalier Universitaire de Nancy, Nancy, France

[13]Cardiovascular and Renal Clinical Trialists, French Clinical Research Infrastructure Network, Nancy, France

[14]Department of Medicine, Bayero University/Aminu Kano Teaching Hospital, Kano, Nigeria

[15]Department of Cardiac Sciences, King Fahad Cardiac Center, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[16]Duke Global Health Institute, Duke University, Durham, NC, United States

[17]Population Health Research Institute, Hamilton Health Sciences, Hamilton, ON, Canada

[18]McMaster University, Hamilton, ON, Canada

Corresponding Author:
Lijing L Yan, MPH, PhD
Global Health Research Center
Duke Kunshan University
No 8 Duke Avenue, Duke Kunshan University, Kunshan, Jiangsu, China
Kunshan, 215316
China
Phone: 86 51236657057
Fax: 86 51257728899
Email: lijing.yan@duke.edu

Abstract

Background: Coronary heart disease (CHD) is the leading cause of cardiovascular mortality worldwide, yet implementation of evidence-based strategies for secondary prevention remains suboptimal.

Objective: This study aimed to evaluate the feasibility, specifically the usability and acceptability, and estimate the preliminary effectiveness of a mobile health (mHealth) intervention targeting both physicians and patients to improve adherence to evidence-based medications and lifestyle modifications.

Methods: We conducted a 12-week pre-post interventional pilot study at two sites in Shanghai and Hainan, China. Physicians used the app designed in this study to prescribe evidence-based medicines and record patient information. Eligible and consenting

patients received automatic text messages or voice calls 4 to 5 times per week for 12 weeks on medication adherence and healthy behaviors. Interviews were conducted among 10 physicians and 24 patients at the two sites for their thoughts on medication adherence and feedback on the usability and acceptability. Questions on usability and acceptability were also asked in a patient follow-up survey. With regard to estimating effectiveness, the primary outcome was medication adherence (as estimated by the Morisky Green Levine Scale) at 12 weeks. Secondary outcomes included physical activity, smoking status, fruits and vegetables consumption, and facility visit frequency.

Results: Interview findings and patient survey showed the good usability and acceptability of the intervention. Among 190 patients who completed the intervention, there was a significant increase in medication adherence (odds ratio [OR] 1.80, 95% CI 1.14-2.85). The study also showed decrease of smokers' percentage (−5%, *P*=.05), increase of daily vegetables consumption frequency (+0.3/day, *P*=.01), and community health care center visit frequency (+3 in 3 months, *P*=.04). The following site-specific differences were noted: medication adherence appeared to increase in Hainan (OR 14.68, 95% CI 5.20-41.45) but not in Shanghai (OR 0.61, 95% CI 0.33-1.12).

Conclusions: Our study demonstrated that the intervention was feasible in both a tertiary care center and an urban community health center in China. Preliminary results from pre-post comparison suggest the possibility that provider and patient-linked mHealth interventions may improve medication adherence and lifestyle modifications among CHD patients, especially in resource-scarce settings. Randomized controlled trials are needed to verify the findings.

(JMIR Mhealth Uhealth 2018;6(1):e9) doi:10.2196/mhealth.7849

KEYWORDS

coronary heart disease; secondary prevention; medication adherence; mobile applications; text messaging

Introduction

Background

Coronary heart disease (CHD) is the leading cause of cardiovascular mortality worldwide [1]. Despite a recent decline in high-income countries [2,3], CHD mortality continues to rise rapidly in low- and middle-income countries [4]. In China, CHD mortality was approximately 100 per 100,000 person in 2013, ranking second only to stroke among causes of cardiovascular deaths [5,6].

Clinical guidelines recommend a combined strategy of using evidence-based medicine and lifestyle modifications for secondary prevention. However, the implementation of these recommendations is suboptimal. The Prospective Urban Rural Epidemiology (PURE) study found that more than 50% of community-based patients from 17 high-, middle-, and low-income countries with known cardiovascular diseases did not take any medicines recommended by guidelines [7]. Even when patients have been initiated on evidence-based medicines at hospital discharge, the medication adherence is usually poor regardless of the socioeconomic status of the patients [8-10]. Evidence from a large multicenter retrospective analysis among more than 2901 CHD patients in China showed that less than 10% of patients used aspirin, clopidogrel, Angiotensin converting enzyme (ACE) inhibitors, or calcium antagonist at 1 year of hospital discharge [8]. Patients were also observed to have a high prevalence of residual lifestyle risk factors [9]. For example, 70% of men and 8% of women with CHD were active smokers in a large Chinese multicenter cross-sectional survey [9].

Addressing the implementation challenges of secondary prevention of CHD requires feasible, scalable, and cost-effective solutions. Recent advances in the widespread use of mobile phones and technology have made mobile health (mHealth) a promising solution. Short message service (SMS) for text messaging is one of the mHealth approaches that has been used in the management of diseases [10] such as asthma [11], human immunodeficiency syndrome (HIV) [12,13], malaria [14], diabetes [15], hypertension [10], and CHD [16,17]. However, the evidence of the effect of mHealth on medication adherence, diet, and physical activities remains inconclusive [10]. Furthermore, few studies have examined multifaceted interventions targeting both health care providers and patients for the improved delivery of secondary prevention of CHD [16].

Information for the Adherence and Knowledge Exchange Heart Disease Medicines Study

We designed and implemented the Adherence and Knowledge Exchange heart disease medicines study (TAKEmeds study) to examine whether an mHealth intervention can improve the patient adherence to evidence-based medications for the secondary prevention of CHD. The intervention includes a provider-facing mobile app guiding medicine prescription and a patient-directed text message or voice call system that promotes medication adherence and behavior modification to optimize secondary prevention of CHD. The specific purpose of this study was to pilot the intervention in China to examine its feasibility, including usability, acceptability, and preliminary effectiveness. This study follows the mHealth evidence reporting and assessment checklist [18].

Methods

Study Design

The TAKEmeds project was a pre-post multicenter interventional pilot study conducted in a community health care center in Shanghai and a tertiary hospital in Hainan Province, China (see Figure 1 for the flowchart of the study). The study was conceived and developed as a part of the World Heart Federation's Emerging Leader program. Implementation in China was led by the Global Health Research Center, Duke Kunshan University, in local partnership with the School of

Public Health, Fudan University, and Hainan Provincial Nongken General Hospital (HPNGH). The study obtained ethical approval from the institutional review boards at Duke University Health System, Fudan University, and HPNGH. The trial was registered in the clinicaltrials.gov database on November 2015 (NCT02597205). The design process of the intervention tools has been previously reported [17].

Study Sites

The TAKEmeds study was conducted in a community health clinic in Longhua Street, Shanghai, and a tertiary care facility in Haikou, Hainan. Shanghai is a large metropolis with around 23 million people and the third highest gross domestic product (GDP) per capita in China, whereas Hainan is a tropical island province in southern China with around 8 million people, ranked 21st out of 34 provinces in terms of GDP per capita [19]. The number of physicians per 1000 population in Shanghai was 2.17 versus 1.38 in Hainan in 2012 [20]. The Longhua Street Community Health Center in Shanghai serves 74,827 populations in 20,789 families. The Hainan Provincial Nongken General Hospital is one of the two tertiary hospitals serving Haikou, the capital city in Hainan, with around 1800 beds serving 590,300 residents. These two sites were selected because of their following distinctive features: primary health care center in a developed area and tertiary hospital in a developing area, which aids generalizability of our findings to disparate health systems across China.

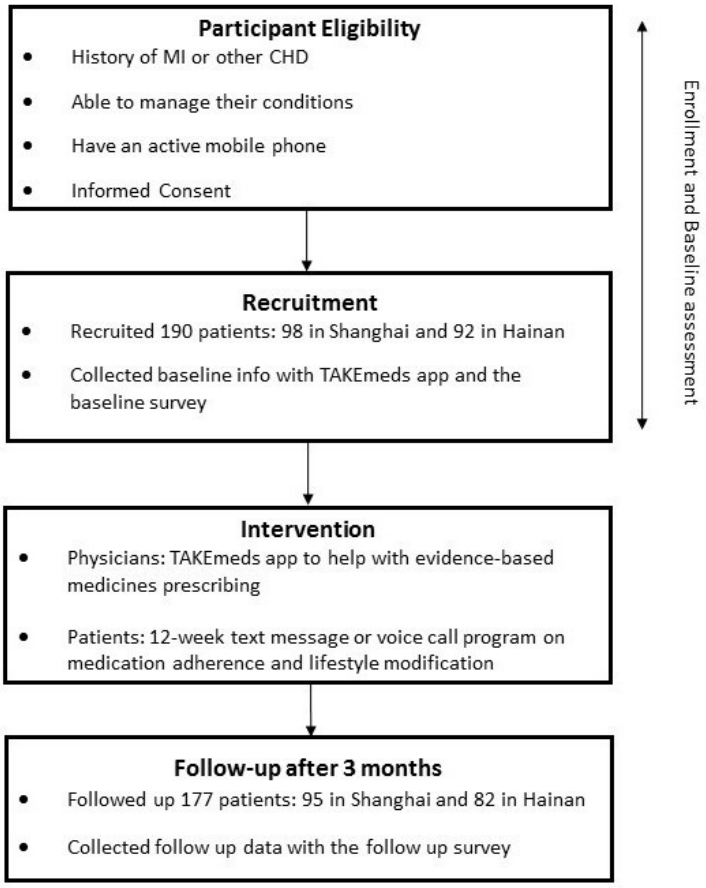
Study Participants

Patients were eligible if (1) they had a history of a myocardial infarction (MI) or obstructive CHD (as clinically diagnosed by the treating physician), (2) they were physically and mentally able to manage their health care themselves, (3) they owned a mobile phone and were comfortable with receiving messages, (4) they were able and willing to provide informed consent, and (5) they were above 18 years of age. Patients who refused consent, participated in another study, or were severally ill with less than 3 months of expected survival were excluded.

In the community health center in Shanghai, patients were identified through the CHD patient list by general practitioners. In the tertiary hospital in Hainan, patients were identified through screening daily admissions and from cardiology outpatient clinic visits. The clinic was affiliated with the tertiary hospital. A promotion flyer that described the study aim and intervention and provided information about how to join the study was distributed to patients for recruitment. Patients were remunerated at the beginning and at end of study with a small gift (combined value 80 Ren Min Bi [RMB] [US $13]).

Physicians were eligible if (1) they owned an Android mobile phone, (2) they took care of coronary artery diseases patients, and (3) they were willing to provide informed consent. Physicians that chose to participate in intervention were given a monetary compensation of around 200 RMB (around US $32) in total.

Figure 1. The Adherence and Knowledge Exchange heart disease medicines (TAKEmeds) study flowchart.



Intervention

The technology-enabled multifaceted intervention was developed to improve patient adherence to medications and modify their lifestyles. At the provider level, we developed an Android-based mobile app, named TAKEmeds, to support prescription of evidence-based medicines and facilitate patient recruitment. At the patient level, we developed a message bank with 60 messages that could be automatically sent to patients through a central server, nonrepetitively, over 12 weeks. These messages were based on current international guideline recommendations and evidence promoting medication adherence and lifestyle modifications. Patients who were active cigarette smokers at the time of enrollment also received tailored messages that help support smoking cessation. The TAKEmeds app and short messages were beta-tested and validated before the study rolled out. The app test version was validated by researchers and physicians, and the messages were tested among the employees in the development company. Details of app and messages development using principles of user-centered design have been described elsewhere [17].

After a brief training session, physicians used the app to record basic demographic and clinical information about eligible patients as well as to prescribe medications for secondary prevention of CHD. The TAKEmeds app made medication recommendations based on the guidelines for secondary prevention of MI published by the UK's National Institute for Health and Care Excellence (NICE) [21]. The NICE guideline was chosen as it was widely accepted and adopted internationally, and there are currently no national guidelines for MI secondary prevention in China. We also consulted a cardiologist in Beijing Anzhen Hospital (one of the coauthors) on its adaption in China before it was adopted. Physicians were able to choose medicines from the recommendations on the app after they entered time interval since heart disease onset (≥or <1 year). The app also provided information for physicians such as an overview of treatment options for MI patients with comorbidities and the importance of lifestyle changes. The flowchart of the app operational procedures and screenshots are presented in Figure 2, and more details can be found in the published development paper elsewhere [17].

At the time of enrollment, patients could choose to receive messages as short text messages or recorded phone calls. The phone calls were provided as an option for senior patients not comfortable with text messages, and they were machine recorded and shared the same contents (details of the messages and phone calls can be found in the paper published elsewhere) [17]. Messages covered the following 5 modules: medication adherence, physical activity, diet, smoking cessation, and general heart health. Example messages of the 5 modules are presented in Figure 3. We set the algorithm to send 4 to 5 messages per week to patients during weekdays for 12 weeks. Nonsmokers received one medication adherence message, one nutrition message, one exercise message, and one general heart health message on 4 random weekdays. Smokers received one additional message on smoking cessations. These messages were sent at one of the four random time slots: 9:00-9:10 AM, 12:00-12:10 PM, 3:00-3:10 PM, or 5:00-5:10 PM on the weekdays. As a part of the enrollment process, patients were also told the messages were one-way only (ie, patients could not respond to the messages with queries), and they could opt to stop receiving messages at any time by either informing their physicians or calling a given number.

Outcome Measures

The primary outcome was change in proportions of patients with different medication adherence status at 12 weeks compared with baseline. Medication adherence was measured by the 4-item Morisky Green Levine Scale [22]. Patients were classified into following 3 categories of adherence: low (score 3-4), intermediate (score 1-2), and high (score 0) [22].

Secondary outcomes included the following: (1) change in proportions of patients with different physical activity level, measured by the short-form version of the International Physical Activity Questionnaire [23]; (2) change in proportions of patients with different smoking status (classified as smoking every day, smoking occasionally, and never smoking); (3) change in the median of frequencies of patients consuming vegetables and fruits, calculated with the 6-item brief dietary assessment tool from the Behavioral Risk Factor Surveillance System fruit and vegetable dietary intake module [24]; and (4) change in the frequencies of patients visiting health facilities over the 3 months.

We conducted a patient survey at the end of the study to assess patient satisfaction toward the intervention, patient perceptions regarding the helpfulness of the intervention, and feedback regarding how they would prefer to receive messages in the future. To assess the affordability of the intervention, we collected basic costs of the messages or voice calls based on the charges of the carrier. We did not collect all costs incurred in this study, as it was not our intention to conduct a full economic analysis.

Sample Size Estimates

Sample size for this pilot study was primarily determined by budget and feasibility. The calculations were also done based on the primary outcome of medical adherence change at 12 weeks. Assuming the odds for patients with high medication adherence versus middle and low medication adherence (or high and middle medication adherence vs low medication adherence) were 10 times larger after the intervention, a total of 74 patients would have 80% power (two-sided alpha of .05). Allowing for 10% loss to follow-up and multiple sites, a total of 166 patients (83 patients in each site) would provide 80% power to detect the difference.

Data Collection and Analysis

Quantitative Data

We collected patient quantitative data from two sources. The first was the self-reported survey at baseline and 12 weeks at follow-up. The survey was administered by trained researchers in Shanghai and resident physicians in Hainan, and they collected data on primary and secondary outcome measures. In Shanghai, trained researchers visited the primary health care centers at baseline and follow-up to collect the data. The second was the TAKEmeds app into which demographic and clinical history were entered by physicians at the time of enrollment.

Figure 2. Flowchart of the Adherence and Knowledge Exchange heart disease medicines (TAKEmeds) app operational procedures.

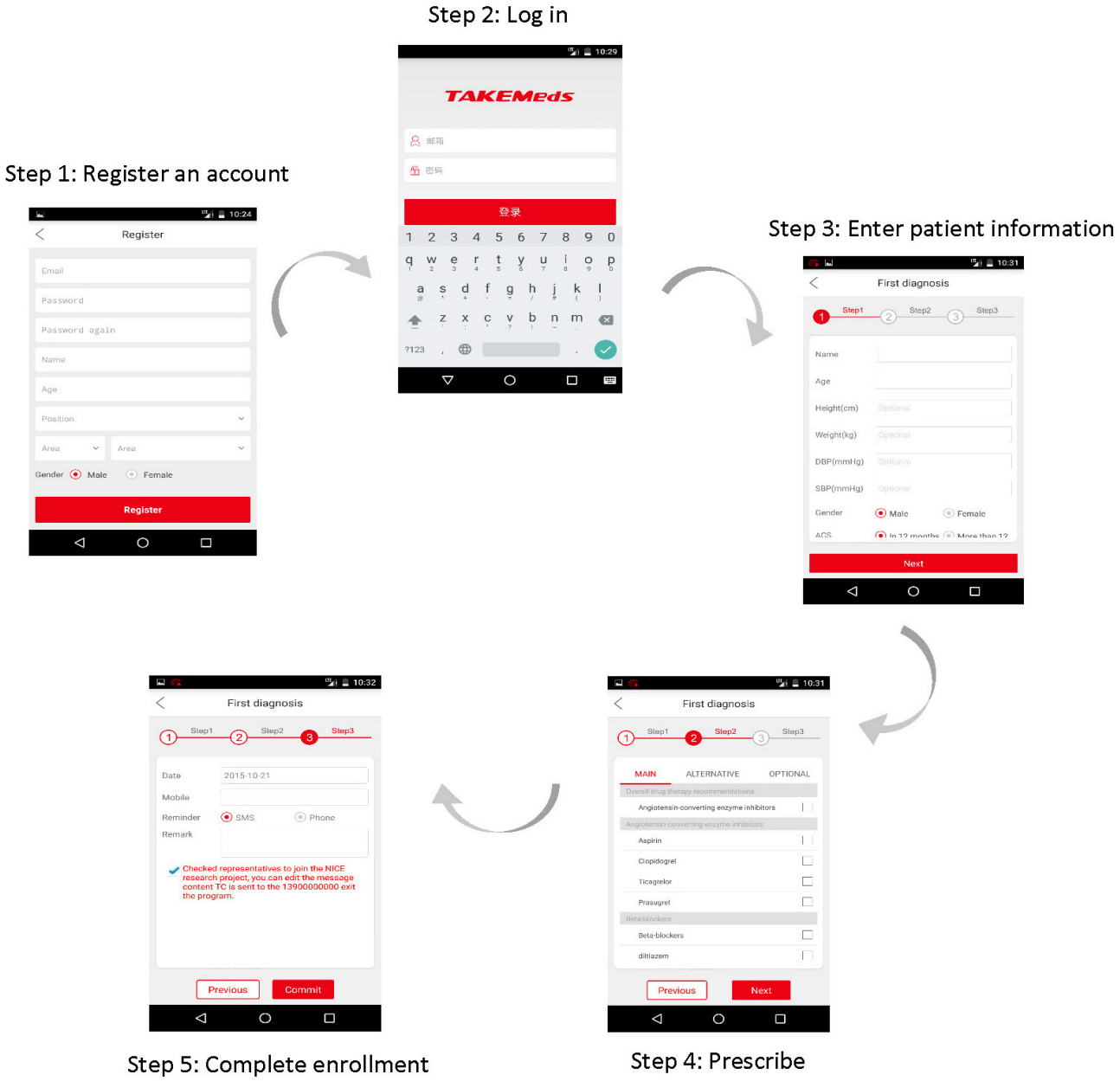


Figure 3. Examples of text messages developed and used in the Adherence and Knowledge Exchange heart disease medicines (TAKEmeds) study.

Smoking
1. It's never too late to stop smoking. Did you know that quitting smoking after a heart attack can significantly improve your health and prevent future heart attacks?
2. Most people who smoke cigarettes, do it in certain situations. Try to be aware and avoid them if you can

Adherence
1. It is important to keep taking your heart medicines even when you are feeling well. They help you maintain your health and live longer.
2. A blood thinner pill, like aspirin helps prevent another heart attack. Ask your doctor about it!

Diet
1. Eating a diet rich in fruits and vegetables along with low-fat dairy products can reduce your blood pressure and maintain your cardiac health.
2. When cooking, try to limit the amount of cooking oil and salt you use.

Exercise
1. Think of fun ways to exercise in your daily routine especially if you are mostly inactive, such as getting off the bus one stop early and walking the rest of the way, practicing Tai Chi, or dancing.
2. After confirming with your physician that your health conditions allow you to have exercise, try to start in a safe way because as little as 30 minutes of brisk walking a day can help lower your cholesterol and blood pressure.

General heart health
1. High blood pressure can damage your heart and your body. Be sure to check your blood pressure frequently.
2. Ask your family and friends for support after a heart attack, they can help you to maintain a healthy lifestyle.

In addition, we estimated communication costs based on the recorded number of messages and the estimated duration of voice calls. The pre-post paired analyses were done at the individual patient level. As for the primary outcome, we built an ordinal (ordered) logistic regression model to test the intervention effect on medication adherence, and both univariate and covariate analyses were conducted. Time (before or after the intervention) was set as the dummy variable. Covariates in the model include age, gender, region, heart disease type, heart disease diagnosis time (≤ or >1 year), hypertension, diabetes status, and number of medications taken. Region was further treated as an interaction term to assess the outcome differences at the two sites. As for secondary outcomes, we used Wilcoxon test to compare the before and after change for physical activity, smoking status, fruits and vegetables consumption, and center visit frequency, and subgroup analysis was conducted. Analyses were conducted using STATA version 13.0 (StataCorp, College Station, TX, USA). All statistical tests were two-tailed with significance level set at .05.

Qualitative Data

The qualitative data were collected through process evaluations at the middle and end of the study. Two focus group discussions were conducted among physicians in Shanghai (n=4) and Hainan (n=6), each lasting around 1 hour. Individual interviews were conducted among randomly selected patients (n=24, 12 from each site) at both sites, each lasting 20 min. The qualitative component was to understand the perceived barriers to improving medication adherence and their feedback on the usability and acceptability of the intervention.

Interviews and focus group discussions were recorded and transcribed. We used the thematic framework to analyze the qualitative results by using the NVivo version 10.0 qualitative data analysis software (QSR International Pty Ltd). Specifically, we coded the transcripts into different nodes of meaning and then grouped the nodes with similar meaning under one theme.

Results

Patient recruitment occurred from May 2015 to August 2015, and follow-up was completed by November 2015. A total of 190 patients were enrolled by 10 physicians, with 98 patients in Shanghai and 92 patients in Hainan. We successfully followed up 177 patients (follow-up rate 93.2%) after the 12-week intervention. Loss to follow-up rate was 3.1% (n=3) in Shanghai and 10.9% (n=10) in Hainan (P=.03). No significant differences were observed between the profiles of the loss-to-follow patients

compared with followed patients at the two sites combined or at each site. The results section presented findings of the preliminary effectiveness, usability, and acceptability of the intervention.

Baseline Characteristics

Baseline demographics, cardiac history, and medications of the 190 patients are presented in Table 1. Patients were 31.6% female, with a mean age of 67 years (SD 10). Patients in Hainan were 5 years younger than those in Shanghai (64 years vs 69 years, *P*=.004) with more severe CHD (a significantly higher proportion had a history of MI, and a higher proportion had been diagnosed with CHD within 1 year). All patients in Hainan took at least 2 cardiac medicines, whereas 79.6% of patients in Shanghai did so. However, fewer patients in Hainan had been previously diagnosed with hypertension (47.8% vs 71.4% in Shanghai). The vast majority of patients (94.2%) chose to receive the information via text messages rather than via phone calls.

The mean age was 62 years (SD 10) for interviewed patients. On average, patients in Hainan (mean age: 57 years, SD 12) were younger than those in Shanghai (mean age: 68 years, SD 8). The percentage of male patients at both sites was 75.0%. Six physicians in Hainan (all cardiologists) and 4 physicians in Shanghai (all general practitioners) participated in this study and focus group discussions. The mean age of physicians was 36 years (SD 5). Physicians in Shanghai (mean age 35 years, SD 4) were generally younger than those in Hainan (mean age: 37 years, SD 6); at both sites, 50.0% of the physicians were males. All Shanghai physicians were in practice for ≤10 years, whereas Hainan physicians were in practice between 11 and 20 years.

Table 1. Demographic, health status, and medicine use of patients at baseline.

Baseline characteristics	Total (N=190)	Shanghai (N=98)	Hainan (N=92)	*P* value
Gender, n (%)				
Female	60 (31.6)	35 (35.7)	25 (27.2)	.21
Male	130 (68.4)	63 (64.3)	67 (72.8)	
Age in years, mean (SD)	67 (10)	69 (8)	64 (12)	.004
Diagnosed with hypertension, n (%)	114 (60.0)	70 (71.4)	44 (47.8)	.001
Diagnosed with diabetes, n (%)	49 (25.8)	28 (28.6)	21 (22.8)	.34
Heart disease type, n (%)				
MI[a]	105 (55.3)	47 (48.0)	58 (63.0)	.03
Non-MI	85 (44.7)	51 (52.0)	34 (37.0)	
Time of diagnosis of heart disease, n (%)				
Within 1 year	67 (35.3)	17 (17.3)	50 (54.3)	<.001
More than 1 year	123 (64.7)	81 (82.7)	42 (45.7)	
Number of cardiac medications, n (%)				
1	20 (10.5)	20 (20.4)	0	
2	37 (19.5)	26 (26.5)	11 (12.0)	<.001
3 and more	133 (70.0)	52 (53.1)	81 (88.0)	
Medication class, n (%)				
ACE[b] inhibitors and ARB[c]	43 (22.6)	27 (27.6)	16 (17.4)	.09
Antiplatelets	152(80.0)	67 (68.4)	85 (92.4)	<.001
Beta-blockers	118 (62.1)	52 (53.1)	66 (71.7)	.01
Aldosterone antagonists	9 (4.7)	0	9 (9.8)	.02
Statins and other lipid-lowering agents	131 (68.9)	59 (60.2)	72 (78.3)	.01

[a]MI: myocardial infarction.

[b]ACE inhibitors: angiotensin-converting enzyme inhibitors.

[c]ARB: angiotensin receptor blockers.

Table 2. Effect of the Adherence and Knowledge Exchange heart disease medicines (TAKEmeds) intervention on the primary outcome: medication adherence.

Medication adherence	Baseline	Follow-up	Unadjusted OR[a] (95% CI)	*P* value	Adjusted[b] OR (95% CI)	*P* value
Combined, n (%)						
High	107 (61.5)	121 (69.5)	1.74 (1.09-2.78)	.02	1.80 (1.14-2.85)	.01
Middle	44 (25.3)	44 (25.3)				
Low	23 (13.2)	9 (5.2)				
Shanghai, n (%)						
High	61 (64.9)	47 (50.0)	0.71 (0.42-1.22)	.22	0.61 (0.33-1.12)	.11
Middle	19 (20.2)	38 (40.4)				
Low	14 (14.9)	9 (9.6)				
Hainan, n (%)						
High	46 (57.5)	74 (92.5)	12.85 (4.62-35.76)	<.001	14.68 (5.20-41.45)	<.001
Middle	25 (31.3)	6 (7.5)				
Low	9 (11.3)	0				

[a]OR: Odds ratio.

[b]Adjusted for age, gender, region, heart disease type, heart disease diagnosis time (< or >1 year), hypertension, diabetes status, and number of medications taken.

Preliminary Estimate of Effectiveness: Primary Outcome

Quantitative Results

The primary outcome results are presented in Table 2. We observed a significant improvement in medication adherence in participants postintervention, either before (odds ratio [OR] 1.74, *P*=.02) or after adjusting for the covariates (OR 1.80, *P*=.01). The percentage of participants reported with high medication adherence increased from 61.5% to 69.5% and with low medication adherence decreased from 13.2% to 5.2%. Participants who took 2 (OR 3.11, *P*=.02) or more than or equal to 3 medicines (OR 2.66, *P*=.03) tended to improve their medication adherence more significantly, compared with those who only took 1 medicine, after adjusting the covariates. Results of the full model can be found in Multimedia Appendix 1.

We identified a significant interaction between location of inclusion and intervention effect on medication adherence (*P*<.001). Results of the model with region as the interaction term can be found in Multimedia Appendix 1. The unadjusted and adjusted outcomes were presented in Table 2 (descriptions here were adjusted outcomes only). In Hainan, there was a significant improvement in medication adherence (OR 14.68, *P*<.001), meaning that after the intervention, the odds for patients in Hainan with high medication adherence versus middle and low medication adherence (or high and middle medication adherence vs low medication adherence) were 14.68 times larger, given the other variables were held constant. This could be shown in the increase of percentage of participants with high medication adherence from 57.5% to 92.5%. In contrast, we did not observe significant change in medication adherence in patients included in Shanghai (OR 0.61, *P*=.11).

Qualitative Results

The qualitative analysis of the patient and physician interviews revealed several factors that might lead to suboptimal medication adherence. Factors arising from the current flawed health system include the following: first, there existed no standard patient follow-up scheme, especially for cardiovascular patients. One physician stated the following:

> *There are follow-up interviews of hypertension and diabetes patients, while none for cardiovascular patients.* [Physician, Shanghai, female, 33 years old]

Another physician stated the following:

> *It is our system's problem. There's no integrated NCD management scheme to follow-up with patients.* [Physician, Hainan, male, 35 years old]

Second, facilities suffered shortage of medicine supplies. One physician stated the following:

> *There are some other medicines. Although they are cheap, and we want to use them, there are no supplies. There's no on the market, and pharmaceutical companies are unwilling to produce*
>
> [Physician, Hainan, female, 38 years old]

Another physician stated the following:

> *It is difficult to prescribemedicines because community hospitals do not have some essential medicines for myocardial infarction patients. We lost them to tertiary hospitals.* [Physician, Shanghai, female, 33 years old]

One physician stated the following:

> *For example, the community hospital only has short-acting Betaloc, instead of long-acting one. And*

> *other medicines are the same. It would be best if I did not have to go to large hospitals, but I have no other way.* [Patient, Shanghai, female, 80 years old]

Third, the intense patient-physician relationship reduced patients' trust in physicians. One physician stated the following:

> *Patients would come only for prescriptions. They would think that community centers do not have good solutions. So it is hard to manage their conditions.* [Physician, Shanghai, male, 30 years old]

Another physician stated the following:

> *There's a crisis of trustiness. There are a lot of people think, we are aiming at something when we ask patient to take medicines.* [Physician, Hainan, female, 38 years old]

Fourth, the unaffordability of the medicines kept patients from adhering to physicians. One physician stated the following:

> *I don't have much income as I have retired. It may be difficult, but there's no stipend for me.* [Patient, Shanghai, male, 68 years old]

Another physician stated the following:

> *The patients may find the medicines are expensive and they cannot afford. So they have to quit it* [Physician, Hainan, female, 38 years old]

Apart from factors attributing to current health system, another reason was that the prevention awareness among the patients was low. One physician stated the following:

> *Patients would only come and see a doctor when symptoms appear.* [Physician, Shanghai, female, 33 years old]

In addition, medicines' side effects were another downside factor for patients to adhere.

Preliminary Estimate of Effectiveness: Secondary Outcomes

Quantitative Results

The secondary outcomes are presented in Table 3. We found an improving trend in patients' smoking status: the percentage of patients that reported no smoking increased from 83.0% to 87.5% (P=.05). We observed an increase in the daily frequency of consuming vegetables (from 2.4 to 2.7/day, P=.01). The results showed an increase in community health care center visit frequency from 7 to 10 during the 12 weeks (P=.04) in Shanghai. We did not find significant changes among patients in terms of fruits consumption (P=.18) or physical activity (P=.91).

Qualitative Results

The interviews and focus group discussions collected physicians' and patients' feedback on whether and how the intervention could help modify lifestyles and enhance physician-patient communication.

Table 3. Paired pre-post comparison of secondary outcomes.

Secondary outcomes	Total			Shanghai			Hainan		
	Baseline	Follow-up	*P* value	Baseline	Follow-up	*P* value	Baseline	Follow-up	*P* value
Physical activity, n (%)									
Highly active[a]	14 (8.0)	17 (9.7)	.91	7 (7.4)	8 (8.4)	.36	7 (8.6)	9 (11.1)	.57
Minimally active[b]	112 (63.6)	105 (59.7)		69 (72.6)	62 (65.3)		43 (53.1)	43 (53.1)	
Inactive[c]	50 (28.4)	54 (30.7)		19 (20.0)	25 (26.3)		31 (38.3)	29 (35.8)	
Current smoking status, n (%)									
Everyday	18 (10.2)	13 (7.4)	.05	8 (8.4)	3 (3.2)	.06	10 (12.3)	10 (12.3)	.36
Seldom	12 (6.8)	9 (5.1)		6 (6.3)	8 (8.4)		6 (7.4)	1 (1.2)	
Never	146 (83.0)	154 (87.5)		81 (85.3)	84 (88.4)		65 (80.2)	70 (86.4)	
Fruits consumption, median frequency/day	0.5	0.6	.18	0.6	0.8	.65	0.3	0.6	.17
Vegetable consumption, median frequency/day	2.4	2.7	.01	2.6	3.2	.001	1.8	2.2	.73
Facility visit, median frequency/3 months	3	3	.22	7	10	.04	1	2	.69

[a]Vigorous-intensity activity on at least 3 days, achieving a minimum of at least 1500 metabolic equivalent (MET)-minutes per week, OR 7 or more days of any combination of walking, moderate-intensity, or vigorous-intensity activities, achieving a minimum of at least 3000 MET-minutes per week.

[b]3 or more days of vigorous activity for at least 20 min per day OR 5 or more days of moderate-intensity activity or walking for at least 30 min per day OR 5 or more days of any combination of walking, moderate-intensity, or vigorous-intensity activities, achieving a minimum of at least 600 MET-min per week.

[c]Individuals who do not meet criteria for the above 2 categories are considered inactive.

The comments we received were generally positive, albeit a few negative reviews remained. One patient stated the following:

> *After I began to receive message, I eat less meat when I was asked, and I work out when I am asked.* [Patient, Shanghai, female, 67 years old]

Another patient stated the following:

> *It helps me receive more knowledge. For example, I received a message asked me to eat less meat. I don't like fish and I only eat meat. And the SMS asked me to eat less of it.* [Patient, Shanghai, male, 74 years old]

One physician claimed that it was hard to change patients' lifestyle:

> *From my perspective, I would say that the effects are few. Because the lifestyle of patients is already mature, it is hard to be changed. And they don't pay much attention to it.* [Physician, Shanghai, male, 33 years old]

Another physician stated that the intervention would take little effect in promoting communication with patients because they were too busy:

> *My patients are too many, and I don't have time to respond at all.* [Physician, Hainan, male, 33 years old]

Usability and Acceptability

All interviewed patients in Shanghai and 10 out of the 12 interviewed patients in Hainan commented that the messages were easy to understand. All participated physicians commented that the app was easy to use; however, the use of prescription function was restricted. It took them 5 to 10 mins to set up a patient on the app.

Table 4. Perceived helpfulness of the intervention program and user feedback.

Characteristics	Total, n (%)	Shanghai, n (%)	Hainan, n (%)
Helpful in			
Medication adherence (n=152)	123 (80.9)	68 (74.7)	55 (90.2)
Healthy diet (n=173)	132 (76.3)	69 (75.8)	63 (76.8)
Exercise (n=173)	115 (66.5)	53 (58.2)	62 (75.6)
Smoking cessation (n=23)	14 (60.9)	3(30.0)	11 (84.6)
Communication with physicians (n=173)	129 (74.6)	59 (64.8)	70 (85.4)
Preferred way of receiving messages (n=171)			
Text messages	111 (64.9)	84 (92.3)	27 (33.8)
Voice calls	57 (33.3)	4 (4.4)	53 (66.3)
Other	3 (1.8)	3 (3.3)	0
Preferred time of receiving messages (n=173)			
8:00-11:59 AM	73 (42.2)	40 (44.0)	33 (40.2)
12:00-12:59 PM	3 (1.7)	2 (2.2)	1 (1.2)
1:00-4:59 PM	14 (8.1)	5 (5.5)	9 (11.0)
5:00-9:00 PM	24 (13.9)	9 (9.9)	15 (18.3)
Does not matter	59 (34.1)	35 (38.5)	24 (29.3)
Preferred frequency of receiving messages (n=171)			
≥Twice/day	3 (1.8)	2 (2.2)	1 (1.3)
Once/day	42 (24.6)	31 (34.1)	11 (13.8)
5-6 times/week	8 (4.7)	6 (6.6)	2 (2.5)
3-4 times/week	17 (9.9)	9 (9.9)	8 (10.0)
1-2 times/week	68 (39.8)	29 (31.9)	39 (48.8)
1-3 times/month	33 (19.3)	14 (15.4)	19 (23.8)
Receiving messages in (n=170)			
One-way only	59 (34.7)	46 (50.5)	13 (16.5)
Two-way	111 (65.3)	45 (49.5)	66 (83.5)
Satisfaction toward the program (n=172)	153 (89.0)	75 (82.4)	78 (96.3)

The overall satisfaction rate with the program was 89.0 % (see Table 4). Participants rated this intervention to be helpful with improving adherence to medications (80.9 %), dietary recommendations (76.3%), increased physical activity (66.5%), and smoking cessation (60.9 %). The majority of patients perceived that the app improved physician-patient communication (74.6%).

We also received suggestions from physicians and patients on how to improve the intervention. Major suggestions included the following: incorporation of the intervention within the current health information system, further customizing message content to individual conditions, and talking slower on the recorded phone call.

A total of 11,534 messages were successfully delivered; however, because of local regulation policy, 166 messages failed to be sent at the beginning stage. All voice calls (720 in total) were successfully dialed and each voice call lasted for 1 to 2 min. During the course of the study, no patient requested cessation of message delivery.

The average cost charged by the carrier was approximately 0.10 RMB (around US $0.015) per text message and 0.09 RMB (around US $0.013) per voice call. To complete the 12-week intervention, it was estimated that for smokers, the average cost was 6 RMB (US $0.9) for the text-message program and 5.4 RMB (US $0.8) for the voice-call program; for nonsmokers, the costs were 4.8 RMB (around US $0.74) and 4.3 RMB (around US $0.66), respectively.

Discussion

Summary of Findings

To our knowledge, this is the first mixed-method study to target both providers and patients in China using mHealth technology to improve the secondary prevention of CHD. The intervention was feasible with easy-to-use products, low operational costs, and well accepted by patients, though the use of app prescription function was challenging for some physicians. We found that a multifaceted intervention that included a provider-facing Android app and a 12-week program of text messages or phone calls directed at patients with CHD was associated with an improvement in patient medication adherence (Hainan), smoking cessation (combined), and vegetable consumption (Shanghai) with notable regional differences that merit additional investigation.

Use of Mobile Health Technology

The ubiquity of mobile phones, especially in resource-limited settings, has made mHealth interventions an attractive option to modify health-related behaviors [25]. Quite a few studies have been conducted to improve the secondary prevention of CHD patients through text messages, and the evidence available seems consistent in supporting the positive effect of the intervention on medication adherence [26-28] and some have demonstrated the effectiveness on smoking cessation [29] and dietary habits [27,28].

In this study, we designed a mobile app for physicians and text messages for patients with the ultimate objective to improve medication adherence. Our working mechanism is to create a new follow-up platform through which patients could receive evidence-based support and recommendations from physicians to increase their compliance and improve their lifestyle.

Interpretation of the Findings

Interestingly, we found an association between the intervention and medication adherence among Hainan patients only. Possible drivers of the site-specific effect were differences in the underlying population, including burden of comorbidities, saturation of secondary prevention knowledge among patients, and hospitalization effect. The greater severity of heart disease conditions among patients recruited in Hainan (reflected in their MI proportion and medications number in Table 1) may account for the bigger changes in medication adherence, given that these patients likely had stronger motivations to follow treatment recommendations. We also observed from the patient interviews that the saturation of secondary knowledge among patients was lower, as they could only access secondary prevention knowledge through hospital physicians and this study. Four patients interviewed in Hainan claimed that this study was their only source for information regarding secondary prevention management. Shanghai patients, in comparison, could access the information through primary care providers, community health lectures, and family members. This would suggest that the benefit of messaging-based interventions might be greatest in low-resource settings with a dearth of reliable sources for health-related information. In addition, patients in Hainan experienced recent hospitalization, which might trigger them to better conform to physicians. Due to lack of control arm, we could not exclude the possibility that it was the recent hospitalization that provoked the change, which needs to be further verified in future studies. Though we found differences in patient age, hypertension history, heart diseases type, and time since diagnosis of heart disease between the two sites, the results suggest that those were not the driving factors of the difference in medication adherence.

The study results showed that the intervention might correlate with the improvement in smoking cessation, which is consistent with other mHealth trials [10,29-31]. We observed the association between intervention and the improvements in dietary habits, though it was unclear why the change of vegetables consumption was obvious in Shanghai only. The visit frequency to the community health center in Shanghai increased from 7 to 10 in 3 months. To regulate overprescribing behaviors of physicians in China, the essential medicines policy restricts that physicians in primary health care centers can only prescribe medicines at 1-month dosage. The increased frequency is a positive sign that the intervention promotes the physician-patient communication, and patients tended to have more frequent visit to the community health care centers.

Barriers Encountered During Implementation

Despite the solid mobile technology infrastructure in China, we encountered some technical barriers at implementation. Our text messages were initially suspended by the carrier for 3 to 5 days after pilot testing, as they were mistaken as fraud or commercial messages because of the high volume of messages. This was quickly rectified after we contacted the customer

service departments of the telecommunications carrier and explained the study protocol. For some patients who chose to receive messages by phone call, the caller ID was not displayed correctly so the patients tended to incorrectly identify them as unsolicited advertising calls and hung up. This issue should be addressed in future studies by assigning a unique caller ID for incoming calls that would be shared with patients at enrollment.

The use of the app prescription function was partially restricted by the current Chinese health system. In the community health center, the function was limited as a result of the newly implemented essential medicine policy. The policy made a list of 307 *essential medicines* that could be stored at primary care centers and sold at low prices in 2009 and updated in 2012 to extend it to 520 medicines (each province can add medicines based on specific contexts) [32,33]. However, it turns out that it indirectly restricts the prescription rights of primary care providers as many truly essential medicines are not on the list. Therefore, some of the medicines suggested by the NICE guidelines, such as clopidogrel, could not be prescribed. Physicians in tertiary hospitals have full prescription rights; however, their heavy workload spared them little time to use the app.

Strengths and Limitations of the Study

Our study shows that the intervention is generally feasible. Our study targets both physicians and patients using technology-enabled intervention. In particular, the use of two disparate test sites—a community health center and a tertiary care hospital—greatly enhance the generalizability of our findings within the Chinese health care system. Differences in the findings from these two sites provided rich insights for future implementation research. Our understanding of the results was enhanced by the use of the mixed-method approach.

A principal limitation is that this study does not have a control group, which increases the risk that the change may not be attributed to the intervention only, although we built a regression model to control for possible confounders to mitigate the risk. Future studies with control arm are needed to evaluate its effectiveness in resource-limited settings. The small sample size and relatively short follow-up may preclude detection of small changes in adherence and lifestyle modifications and evaluation of persistence of efficacy over time. The self-reported nature of the outcomes means that results may be contaminated by recall or social likeability biases. Future studies may consider more objective measures, including pill counting for medication adherence, physiologic assessments (blood pressure, cholesterol, and glucose control), and clinical outcomes (MI, revascularization, and death). It may be necessary to include additional confounding variables, such as socioeconomic and psychosocial status, to sufficiently demonstrate the effects. Moreover, future studies should further examine fidelity to the intervention, that is, how many people actually read the messages or listened to the calls to accurately reflect their adherence to the intervention. Future studies could also record data related to number of eligible patients and reasons for nonparticipation to enhance the transparency of the recruitment process. Finally, the intervention was not incorporated within the current workflow of Chinese physicians, which was a barrier to its rapid and widespread adoption. Despite these limitations, this study provides experiences of implementing the TAKEmeds intervention model, its feasibility, and preliminary effectiveness results in two different settings in China.

Conclusions

Both international and national guidelines targeting medicine use and lifestyle modifications are well established to improve the secondary prevention of CHD; however, its uptake into clinical practice is far from optimal in China and around the world.

Such challenges require innovative, feasible, and cost-effective solutions. Harnessing the ubiquity of mobile phones and rapid advance in low-cost mobile technology, this study devised a multifaceted intervention package that targets both providers and physicians to increase the uptake of evidence-based secondary prevention of CHD in China. The results support the feasibility of the intervention: easy to use; well accepted by patients; and showing potential effect in improving medication adherence, smoking status, and possibly vegetables consumption as well as physician-patient interaction in community health centers, though we believe future studies with control group are needed to verify it before scaling up this intervention. We further found that the TAKEmeds intervention might have greater potential to improve the outcomes in resource-limited places such as Hainan, and further studies with control arms should verify the impact of the mHealth intervention across various economical settings. Although this study has several limitations, it provides a proof of concept for the role of mHealth in cardiovascular prevention. If these findings are confirmed in future studies with control arm and clinical endpoints, the intervention can be scaled up across the health system in China and other low-resource settings and can be adapted to other chronic diseases such as hypertension and diabetes.

Acknowledgments

The authors would like to acknowledge their academic partners including McMaster University and the Population Health Research Institute, Canada, which were involved in the study throughout the entire stage, and School of Public Health, Fudan University, which was in charge of the study implementation in Shanghai. The authors would also like to thank the Department of Cardiology, Hainan Nongken General Hospital, Haikou, Hainan Province, for their support of its implementation in Hainan. Specifically, the authors would like to thank the physicians in the Longhua Street Community Health Center and the physicians and residents in Hainan Nongken General Hospital that participated in and contributed to this study. The authors would also like to thank the Zhengjiang Innovation Zhejiang University Kunshan Innovation Institute for their help with app and message development. In addition, the authors acknowledge the contribution of Mr Yuliang Wang from School of Public Health, Fudan University; Dr Ye Sun from Hainan Nongken General Hospital for their research assistance and coordination; Prof Keith Dear and Dr Chaoyun Li

from Duke Kunshan University; Ms Zhe Xu from John Hopkins University and Ms Yechu Hua from Duke University for their statistical support; Ms Minhui Yang for her qualitative research analysis support; Mr Heng Jiang for his help with references; and Louisa Hong and Kelly Sun from McMaster University for their support to develop the messages. Finally, the authors would like to thank World Heart Federation Emerging Leader Seed and the Duke Kunshan University Global Health Master Student Fieldwork for their funding support to this study. Travel support from the institution of each emerging leaders to attend the World Heart Federation (WHF) conference is also appreciated.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental file (tables).

[PDF File (Adobe PDF File), 21KB - mhealth_v6i1e9_app1.pdf]

References

1. Zhou M, Wang H, Zhu J, Chen W, Wang L, Liu S, et al. Cause-specific mortality for 240 causes in China during 1990-2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet 2016 Jan 16;387(10015):251-272. [doi: 10.1016/S0140-6736(15)00551-6]
2. Writing Group Members, Mozaffarian D, Benjamin E, Go AS, Arnett DK, Blaha MJ, American Heart Association Statistics Committee, Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association. Circulation 2016 Jan 26;133(4):e38-360. [doi: 10.1161/CIR.0000000000000350]
3. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe 2014: epidemiological update. Eur Heart J 2014 Nov 7;35(42):2950-2959. [doi: 10.1093/eurheartj/ehu299]
4. Reddy KS. Cardiovascular disease in non-Western countries. N Engl J Med 2004 Jun 10;350(24):2438-2440. [doi: 10.1056/NEJMp048024] [Medline: 15190135]
5. Zhang XH, Lu ZL, Liu L. Coronary heart disease in China. Heart 2008 Sep;94(9):1126-1131. [doi: 10.1136/hrt.2007.132423] [Medline: 18703693]
6. Jiang G, Wang D, Li W, Pan Y, Zheng W, Zhang HY, et al. Coronary heart disease mortality in China: age, gender, and urban-rural gaps during epidemiological transition. Rev Panam Salud Publica 2012 Apr;31(4):317-324. [Medline: 22652972]
7. Yusuf S, Islam S, Chow CK, Rangarajan S, Dagenais G, Diaz R, Prospective Urban Rural Epidemiology (PURE) Study Investigators. Use of secondary prevention drugs for cardiovascular disease in the community in high-income, middle-income, and low-income countries (the PURE Study): a prospective epidemiological survey. Lancet 2011 Oct 1;378(9798):1231-1243. [doi: 10.1016/S0140-6736(11)61215-4]
8. Bi Y, Gao R, Patel A, Su S, Gao W, Hu D, CPACS Investigators. Evidence-based medication use among Chinese patients with acute coronary syndromes at the time of hospital discharge and 1 year after hospitalization: results from the Clinical Pathways for Acute Coronary Syndromes in China (CPACS) study. Am Heart J 2009 Mar;157(3):509-516.e1. [doi: 10.1016/j.ahj.2008.09.026]
9. Liu H, Yu J, Pan C, Chen F, Li J, Hu D. [Cross-sectional study on the prevalence of metabolic syndrome in patients with coronary heart disease in China]. Zhonghua Yi Xue Za Zhi 2006 Aug 15;86(30):2095-2098. [Medline: 17064610]
10. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12. [doi: 10.1002/14651858.CD007459.pub2]
11. Strandbygaard U, Thomsen SF, Backer V. A daily SMS reminder increases adherence to asthma treatment: a three-month follow-up study. Respir Med 2010 Feb;104(2):166-171. [doi: 10.1016/j.rmed.2009.10.003]
12. Mbuagbaw L, Thabane L, Ongolo-Zogo P, Lester RT, Mills E, Volmink J, et al. The Cameroon mobile phone SMS (CAMPS) trial: a protocol for a randomized controlled trial of mobile phone text messaging versus usual care for improving adherence to highly active anti-retroviral therapy. Trials 2011 Jan 7;12:5. [doi: 10.1186/1745-6215-12-5]
13. Lester RT, Ritvo P, Mills EJ, Kariri A, Karanja S, Chung MH, et al. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. Lancet 2010 Nov 27;376(9755):1838-1845. [doi: 10.1016/S0140-6736(10)61997-6]
14. Ollivier L, Romand O, Marimoutou C, Michel R, Pognant C, Todesco A, et al. Use of short message service (SMS) to improve malaria chemoprophylaxis compliance after returning from a malaria endemic area. Malar J 2009 Oct 23;8:236. [Medline: 10.1186/1475-2875-8-236]
15. Kim HS, Kim NC, Ahn SH. Impact of a nurse short message service intervention for patients with diabetes. J Nurs Care Qual 2006;21(3):266-271. [Medline: 16816608]
16. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017]

17. Chen S, Gong E, Kazi DS, Gates AB, Karaye KM, Girerd N, et al. Development of a mobile phone-based intervention to improve adherence to secondary prevention of coronary heart disease in China. J Med Eng Technol 2016;40(7-8):372-382. [doi: 10.1080/03091902.2016.1213904]
18. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174]
19. China Statistical Year Book. China: National Bureau of Statistics of China; 2015.
20. China Health Statistical Yearbook. China: National Bureau of Statistics of China; 2013.
21. National Institute for Health and Care Excellence. telfordccg. 2013. Myocardial infarction: Secondary prevention pathway URL: http://www.telfordccg.nhs.uk/your-health/medicines-management/prescribing-guidelines/cardiovascular/myocardial-infarction/2339-secondary-prevention-myocardial-infarction-nice-pathway/file [accessed 2017-12-22] [WebCite Cache ID 6vttCpqId]
22. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med care 1986 Jan;24(1):67-74. [Medline: 3945130]
23. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [Medline: 12900694]
24. CDC. 2015. Surveillance of Fruit and Vegetable Intake Using the Behavioral Risk Factor Surveillance System URL: https://www.cdc.gov/brfss/pdf/fruits_vegetables.pdf [accessed 2017-12-18] [WebCite Cache ID 6vnf9mtIH]
25. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040]
26. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027]
27. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and Internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237. [doi: 10.2196/jmir.4944]
28. Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779. [doi: 10.1136/heartjnl-2014-305783]
29. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945]
30. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jun 30;378(9785):49-55. [doi: 10.1016/S0140-6736(11)60701-0]
31. Whittaker R, Borland R, Bullen C, Lin RB, McRobbie H, Rodgers A. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2009 Oct 07(4):CD006611. [doi: 10.1002/14651858.CD006611.pub2] [Medline: 19821377]
32. Ministry of Health of the People's Republic of China. moh.gov. 2009. Catalog of National Essential Medicines (Provision of Use for Primary Health Care Institutions) URL: http://www.moh.gov.cn/mohzcfgs/pgz/200908/42506.shtml [accessed 2017-12-26] [WebCite Cache ID 6vzkklCfK]
33. People's Republic of China Ministry of Health Order. SDA. 2012. National Essential Medicines List URL: http://www.sda.gov.cn/WS01/CL0053/79110.html [accessed 2017-12-26] [WebCite Cache ID 6vzkWuPJe]

Abbreviations

ACE: angiotensin-converting enzyme inhibitors.
ARB: angiotensin receptor blockers.
CHD: coronary heart disease
GDP: gross domestic product
HPNGH: Hainan Provincial Nongken General Hospital
MET: Metabolic Equivalent
mHealth: mobile health
MI: myocardial infarction
NICE: National Institute for Health and Care Excellence
OR: odds ratio
RMB: Ren Min Bi (Chinese Currency)
SMS: short message service
TAKEmeds: the Adherence and Knowledge Exchange heart disease medicines study

Edited by G Eysenbach; submitted 11.04.17; peer-reviewed by M Tian, L Park, A Salzwedel; comments to author 22.07.17; revised version received 20.09.17; accepted 30.10.17; published 25.01.18.

Please cite as:
Chen S, Gong E, Kazi DS, Gates AB, Bai R, Fu H, Peng W, De La Cruz G, Chen L, Liu X, Su Q, Girerd N, Karaye KM, Alhabib KF, Yan LL, Schwalm JD
Using Mobile Health Intervention to Improve Secondary Prevention of Coronary Heart Diseases in China: Mixed-Methods Feasibility Study
JMIR Mhealth Uhealth 2018;6(1):e9
URL: http://mhealth.jmir.org/2018/1/e9/
doi:10.2196/mhealth.7849
PMID:29371178

Original Paper

A Smartphone App (BlueIce) for Young People Who Self-Harm: Open Phase 1 Pre-Post Trial

Paul Stallard[1,2], PhD; Joanna Porter[2], BSc (Hons); Rebecca Grist[1], PhD

[1]Child and Adolescent Mental Health Group, Department for Health, University of Bath, Bath, United Kingdom

[2]Child and Family Mental Health,, Temple House, Oxford Health National Health Service Foundation Trust, Keynsham, United Kingdom

Corresponding Author:
Paul Stallard, PhD
Child and Adolescent Mental Health Group
Department for Health
University of Bath
Wessex House 6.10
Claverton Down
Bath, BA2 7AY
United Kingdom
Phone: 44 1225 383282
Fax: 44 1225 383282
Email: p.stallard@bath.ac.uk

Abstract

Background: Recent years have seen a significant increase in the availability of smartphone apps for mental health problems. Despite their proliferation, few apps have been specifically developed for young people, and almost none have been subject to any form of evaluation.

Objective: This study aimed to undertake a preliminary evaluation of a smartphone app (BlueIce), coproduced with young people and designed to help young people manage distress and urges to self-harm. We aimed to assess the acceptability, safety, and use of BlueIce and to explore the effects on the primary outcome of self-harm and the secondary outcomes of psychological functioning.

Methods: We undertook an open trial where we recruited young people aged 12 to 17 years attending specialist child and adolescent mental health services (CAMHS) who were currently self-harming or had a history of self-harm. Eligible participants were assessed at baseline and then given BlueIce. They were assessed 2 weeks later (post familiarization) and again at 12 weeks (post use). A behavior-screening questionnaire (Strengths and Difficulties Questionnaire) was completed along with standardized measures of depression (Mood and Feelings Questionnaire or MFQ) and anxiety (Revised Child Anxiety and Depression Scale or RCADS), taking into account self-reports of self-harm, app helpfulness, and safety.

Results: All core CAMHS professional groups referred at least 1 young person. Out of 40 young people recruited, 37 (93%) elected to use BlueIce after familiarization, with 29 out of 33 (88%) wanting to keep it at the end of the study. No young person called the emergency numbers during the 12-week trial, and no one was withdrawn by his or her clinician due to increased risk of suicide. Almost three-quarters (73%) of those who had recently self-harmed reported reductions in self-harm after using BlueIce for 12 weeks. There was a statistically significant mean difference of 4.91 (t_{31}=2.11; P=.04; 95% CI 0.17-9.64) on postuse symptoms of depression (MFQ) and 13.53 on symptoms of anxiety (RCADS) (t_{30}=3.76; P=.001; 95% CI 6.17-20.90), which was evident across all anxiety subscales. Ratings of app acceptability and usefulness were high.

Conclusions: Our study has a number of methodological limitations, particularly the absence of a comparison group and a prospective way of assessing self-harm. Nonetheless, our findings are encouraging and suggest that BlueIce, used alongside a traditional CAMHS face-to-face intervention, can help young people manage their emotional distress and urges to self-harm.

(JMIR Mhealth Uhealth 2018;6(1):e32) doi:10.2196/mhealth.8917

KEYWORDS

self-injury; smartphone; mobile apps; BlueIce; adolescents; cognitive behavioral therapy; mHealth

Introduction

Self-Harm

Self-harm has been defined as the intentional self-poisoning or self-injury, irrespective of the type of motive or the extent of suicidal intent [1]. Self-harm in young people aged up to 18 years is a major public health problem [2] with community surveys reporting that up to 20% of young people will self-harm by the age of 18 years [3-7]. Figures from hospital episode statistics show that in 2014-2015, 16,000 young people attended accident and emergency department (A&E) in England following self-poisoning [8], whereas a survey of National Health Service (NHS) trusts found that 19,000 adolescents were hospitalized in England and Wales following self-harm in 2016 [9].This statistic underestimates the scale of the problem because the majority of self-harm remains hidden with comparatively few episodes being treated in a hospital [10].

Although self-harm in young people is common, few intervention studies have been reported [11]. Evidence that is available suggests that therapeutic assessment, mentalization, dialectical behavioral therapy (DBT), and cognitive-behavioral therapy (CBT) may be promising and warrant further evaluation [11]. Key challenges for future research are the development and assessment of innovative interventions that are acceptable to young people, which reduce the risk of self-harm and enhance meaningful engagement with health services [12].

Digital Health

In the United Kingdom, the NHS is being encouraged to harness digital technology to enhance and support psychological health [13]. Digital health technology may be particularly attractive for young people who are very familiar with and frequent users of digital media. It offers an accessible way of supporting health services because nearly all young people aged 12-15 years have Internet access and 90% own a smartphone [14].

One form of digital technology, smartphone apps, has seen particular growth [15]. However, few apps have been specifically developed for young people with mental health problems, almost none have been subject to any form of evaluation, and none have been developed specifically for children who self-harm [16]. It is essential that smartphone apps for adolescents are subject to research evaluation [16] and codesigned with people who have lived experience [11].

We developed and coproduced a smartphone app (BlueIce) with young people with lived experience of self-harm. BlueIce is to be used in conjunction with face-to-face interventions and provides a personalized toolbox of strategies founded on evidence-based approaches that can be accessed at any time. The aim of this study was to explore the safety, acceptability, feasibility, and usability of a novel smartphone app, BlueIce, developed specifically for young people who self-harm. We will explore the effect on self-harm (primary outcome) and also on psychological functioning (secondary outcome).

Methods

BlueIce

We coproduced a smartphone app, BlueIce, with young people with a lived experience of self-harm. We adopted a user-centered, agile development process to create, refine, and evaluate BlueIce. This was a tripartite process whereby the product users (young people and clinical staff), academics, and app developers worked together. This process aimed to maximize the acceptability and use of the app by our target group and ensured that the content reflected current evidence and best clinical practice. Informed by the available evidence, BlueIce provides a personalized toolbox of strategies based on CBT and DBT that can be accessed 24/7 [1,11]. It includes a mood diary, menu of personalized mood-lifting activities, and automatic routing through safety checks to delay or prevent self-harm. Mood-lifting activities are designed to improve mood and include a personalized music library of uplifting music, photo library of positive memories, physical activities, mood-changing activities, audio-taped relaxation and mindfulness exercises, identification and challenging of negative thoughts, a contact list of key people to call or text, and distress tolerance activities (informed by DBT). After using the mood-lifting section, the young person is asked to rerate his or her mood, and if the urge to self-harm has not reduced, that young person is automatically routed to emergency numbers (nominated contact, Childline, 111) that he or she can call.

BlueIce is for young people aged 12-17 years and is designed as an adjunct to face-to-face therapy. It is available for use on Apple and Android devices. It has met the minimum standards required for NHS-accredited apps [17]. BlueIce is password protected and all data are stored locally on the device. BlueIce is a prescribed app and is available via license to child and adolescent mental health services (CAMHS). Information about the app is available at the webpage)[18].

Study Design

This is a phase 1 open uncontrolled trial. A detailed summary of the methodology is provided in the study protocol [19].

Setting

Young people were recruited from specialist CAMHS outpatient clinics provided by Oxford Health NHS Foundation Trust. The trust provides CAMHS across Bath and North East Somerset, Buckinghamshire, Oxfordshire, Swindon, and Wiltshire in the United Kingdom.

Participants

Eligible young people aged 12-17 years who were currently self-harming or had a history of self-harm were included. Young people were excluded if first, they were seriously contemplating or planning a suicide attempt. Given that we do not know whether BlueIce will have any unintentional adverse effects, it would not be safe to test it with a high-risk group who were actively suicidal. Second, young people were excluded if they were diagnosed with psychosis or had a significant learning disability, which might impede their ability to use the app. Third, those who had been subject to abuse within the last 6 months

or were the subject of a safeguarding investigation were excluded. Finally, BlueIce is only available in English, and we therefore excluded those who were not proficient in English.

Recruitment and Consent

Eligible participants were identified by their CAMHS clinician and provided with a project information sheet. Interested young people, and their parents if they are under 16 years, met with a researcher to get familiarized with the project. Signed consent was obtained from young people above the age of 16 years, and signed parent consent and child assent from those under 16 years. The project was reviewed and approved by the South West—Exeter Research Ethics Committee (16/SW/0018).

Consenting participants were provided with BlueIce but continued to attend face-to-face meetings with their CAMHS clinician. As is usual practice, the CAMHS clinician was responsible for reviewing risk and implementing any risk plans required to keep the young person safe.

Assessments

Standardized questionnaires were completed at baseline, post familiarization (2 weeks after using BlueIce), and post use (12 weeks).

Depression

Symptoms of depression were assessed by the Mood and Feelings Questionnaire (MFQ), a self-report questionnaire for depression [20]. The MFQ consists of 33 items, which the young person rates as either *true* (scores 2), *sometimes true* (scores 1) or *not true* (scores 0). The MFQ has high criterion validity and correlates well with other measures of depression [21]. A total score of 27 and above is associated with major depression, 20 with mild depression, and 16 with no mood disorder [21-23].

Anxiety

Symptoms of anxiety were assessed by the Revised Child Anxiety and Depression Scale (RCADS) [24]. RCADS is a 47-item questionnaire with items corresponding to Diagnostic and Statistical Manual 4th Edition (DSM-IV) criteria for anxiety in the areas of social phobia, separation anxiety, obsessive compulsive disorder, panic disorder, and generalized anxiety disorder and also for other major depressive disorders. Each item is rated on a 4-point Likert scale of frequency ranging from *never* (0) to *always* (3), and items are then summed to produce subscale and total anxiety scores. There are age- and gender-related norms for identifying clinically significant scores (total score≥64-80) [24]. If the young person was under 16 years, his or her parents were also asked to complete this measure.

Behavior

The Strengths and Difficulties Questionnaire (SDQ) is a brief, widely used behavioral screening questionnaire [25]. The SDQ consists of 25 items that assess emotional symptoms, conduct problems, hyperactivity and/or inattention, peer relationship problems, and prosocial behavior. Participants also rate overall distress and social impact of their behavior on home life, friendships, classroom learning, and leisure activities. Parents also completed this measure if the young person was under 16 years.

Safety

At post familiarization, young people were asked whether BlueIce makes them feel like harming themselves even more, whether they felt able to use BlueIce if they had thoughts of self-harm, and whether they thought BlueIce would help them stop self-harming. Participants rated their responses on a 5-point Likert scale ranging from 1 (*definite no*) to 5 (*definitely*).

Acceptability

At post use, young people used a 10-point Likert scale to rate the ease of use, helpfulness, and whether they would recommend BlueIce to a friend.

Self-Harm

Information on baseline levels of self-harm was obtained from clinical records spanning the 4 weeks before using BlueIce. Self-harm during the trial was assessed via self-report during the interview at baseline and 12 weeks. By estimating the number of self-harm acts that would have occurred if participants continued to self-harm at the same rate through the 12-week trial, we calculated the number of self-harm acts prevented for the young people who (1) stopped self-harming at post use and (2) continued to self-harm at post use but at a reduced rate.

Results

Demographics and Baseline Assessments

Between May and October 2016, 37 different clinicians from 8 CAMHS teams referred 54 young people. All core professionals that constitute CAMHS, that is, child psychiatrists, clinical psychologists, family therapists, child psychotherapists, occupational therapists, and community psychiatric nurses, referred at least 1 young person. Of the 54 referrals, 4 did not meet inclusion criteria, 3 we were unable to contact, 2 dropped out of CAMHS before they could consent, and 1 declined to participate. The remaining 44 completed baseline assessments.

Participants were predominantly girls (40, 91%) with an average age of 16.0 years (SD=1.4), with 30 (68%) having self-harmed at least once in the 4 weeks before starting the trial.

Using recommended cut-offs, 42 young people out of 44 (96%) scored 29 or more on the MFQ, suggesting probable depression. Using age- and gender-adjusted cut-offs on the RCADS, 37 out of 44 (84%) screened positive for one or more anxiety disorders and 37 out of 44 (84%) scored above cut-offs on the SDQ for a probable emotional disorder. On the RCADS, 16 out of 17 (94%) parents rated their child above the cut-off for depression and 16 out of 18 parents (89%) scored their child above the cut-off on the SDQ for significant emotional problems. Demographic data and baseline psychopathology are summarized in Tables 1 and 2.

Data summarized in Table 2 indicate that both young people and parents identified significant problems (young people=85%; parents=95%), which had been present for longer than 12 months (young people=85%; parents=95%) and caused

significant distress (young people=78%; parents=100%). The majority of young people reported significant impairment in friendships (80%), ability to learn (76%), leisure activities (73%), and home life (61%).

Post Familiarization

Postfamiliarization assessments were completed with 40 of the 44 (91%) participants, and of these, 37 (93%) elected to use BlueIce. Those who did not want to use BlueIce reported that they were not ready to stop self-harming. Table 3 summarizes postfamiliarization data.

No safety issues were identified and there were no unintended negative effects on self-harm. No young person thought that BlueIce would increase his or her self-harm with 32 out of 40 (80%) strongly endorsing this statement.

In total, 33 out of 40 (83%) thought they would be able to use BlueIce if they had thoughts of self-harm. Only 2 (5%) felt they would be unable to do so feeling that their urges to self-harm would be too powerful to resist. Young people were less sure about whether BlueIce would help them to stop self-harming, 24 (60%) thought it would, 8 (20%) were unsure, and 6 (15%) felt it might, with 2 (5%) feeling it would not.

Post Use

Postuse assessments were completed with 33 out of the 37 (89%) young people, with paired data being available for 30-32 young people according the specific items analyzed. Data are summarized in Table 4.

Table 1. Characteristics of participants.

Characteristic	Number of participants, n
Gender	
Male	4
Female	40
Age (years)	
12	1
13	5
14	3
15	9
16	15
17	11
Self-harm	
Self-harmed at least once in past 4 weeks	30
No self-harm in last 4 weeks	14

Table 2. Young person and parent-rated baseline symptomatology.

Measure	Child report	Parent report
Mood and Feelings Questionnaire, mean (SD)	43.6 (9.6)	N/A[a]
Revised Child Anxiety and Depression Scale, mean (SD)	81.0 (21.9)	75.4 (26.9)
Strengths and Difficulties Questionnaire, mean (SD)	21.4 (3.3)	23.8 (6.3)
Definite or severe problem, n (%)	33 (85)	17 (95)
Problem present longer than 12 months, n (%)	33 (85)	17 (95)
Causes medium to great deal of distress, n (%)	31 (78)	18 (100)
Effect on home life (medium to great deal), n (%)	23 (61)	14 (78)
Effect on friendships (medium to great deal), n (%)	31 (80)	16 (89)
Effect on ability to learn (medium to great deal), n (%)	28 (76)	11 (61)
Effect on leisure (medium to great deal), n (%)	29 (73)	14 (78)
Burden on you and family (quite a lot to great deal), n (%)	31 (80)	15 (83)

[a]N/A: not applicable.

Table 3. BlueIce post familiarization (2 weeks).

Question	No	Maybe	Not sure	Think so	Definitely
Would you be able to use BlueIce if you had thoughts of self-harm?	2	2	3	20	13
Do you think BlueIce might make you harm more?	32	4	4	0	0
Do you think BlueIce would help you to stop harming?	2	6	8	15	9

Table 4. Paired baseline and follow-up scores on standardized measures.

Measure	Baseline, mean (SD)	Follow-up, mean (SD)	Significance	
			t (degrees of freedom)	*P* value
Self-report (n=30-32)				
Mood and Feelings Questionnaire (MFQ)	42.75 (10.73)	37.84 (15.44)	2.11 (31)	.04
Revised Child Anxiety and Depression Scale (RCADS)				
Panic disorder	14.00 (7.31)	11.20 (6.40)	2.90 (29)	.007
Separation anxiety disorder	8.90 (4.20)	7.37 (4.97)	2.77 (28)	.01
Generalized anxiety disorder	11.27 (3.50)	9.50 (4.05)	2.72 (29)	.01
Social anxiety disorder	19.67 (5.65)	16.60 (6.33)	3.58 (29)	.001
Obsessive compulsive disorder	6.97 (4.21)	5.70 (4.74)	2.64(29)	.01
Depression	19.16 (5.13)	16.58 (6.62)	2.40 (30)	.02
Total RCADS	80.33 (23.75)	66.80 (28.46)	3.76 (29)	.001
Strengths and Difficulties Questionnaire (SDQ)				
Hyperactivity scale	5.44 (1.63)	5.22 (1.83)	0.62 (31)	.54
Emotional symptoms scale	7.91 (1.51)	7.06 (2.17)	2.90 (31)	.007
Peer problems scale	4.91 (1.55)	5.25 (1.61)	−1.36 (31)	.18
Prosocial scale	7.34 (2.24)	7.63 (1.56)	−1.01 (31)	.32
Conduct problems scale	2.91 (1.23)	2.75 (1.05)	0.67 (31)	.51
Total SDQ	21.16 (3.35)	20.28 (4.47)	1.16 (31)	.26
Parent-report (n=10-13)				
Revised Child Anxiety and Depression Scale (RCADS)				
Panic disorder	13.70 (6.36)	12.10 (4.04)	1.50 (9)	.17
Separation anxiety disorder	10.73 (5.29)	9.91 (4.74)	1.22 (10)	.25
Generalized anxiety disorder	11.55 (4.74)	10.00 (2.93)	1.76 (10)	.11
Social anxiety disorder	18.64 (5.56)	17.64 (6.14)	1.08 (10)	.31
Obsessive Compulsive Disorder	9.10 (5.86)	9.50 (5.15)	0.58 (9)	.59
Depression	17.91 (6.49)	17.64 (6.35)	0.27 (10)	.79
Total RCADS	80.90 (30.49)	77.70 (23.21)	1.03 (9)	.33
Strengths and Difficulties Questionnaire (SDQ)				
Hyperactivity scale	4.31 (1.65)	4.15 (1.91)	0.41 (12)	.69
Emotional symptoms scale	8.17 (2.37)	7.25 (2.60)	2.42 (11)	.03
Peer problems scale	8.38 (3.82)	5.77 (1.79)	2.50 (12)	.03
Prosocial scale	7.23 (1.64)	6.54 (1.85)	1.03 (12)	.32
Conduct problems scale	2.92 (1.12)	3.23 (1.83)	0.72 (12)	.49
Total SDQ	24.58 (6.95)	20.50 (6.25)	4.67 (11)	.001

Table 5. Self-harm status at baseline and post use.

Self-harm status	Number of participants, n
At baseline	
Not self-harmed in 4 weeks before baseline assessment	7
Self-harmed in 4 weeks before baseline assessment	26
Post use (12-week)	
Not self-harmed at follow-up	7
Self-harmed during follow-up	0
Not self-harmed during follow-up	4
Self-harmed during follow-up at a reduced rate	15
Self-harmed during follow-up at same rate	7

Depression, Anxiety, and Behavior

At follow-up, there was a statistically significant mean difference of 4.91 (t_{31}=2.11; *P*=.04; 95% CI 0.17-9.64) on postuse symptoms of depression (MFQ) and 13.53 on symptoms of anxiety (RCADS) (t_{30}=3.76; *P*=.001; 95% CI 6.17-20.90), which was evident across all anxiety subscales. There was no statistically significant change on the behavior-screening questionnaire (SDQ) other than a significant mean difference of 0.84 on the emotional subscale (t_{29}=2 .90; *P*=.007; 95% CI 0.25-1.44).

The analysis was repeated comparing those who reported no change in self-harm (n=7) with those who had not harmed or did so at a reduced rate (n=26). There were no postuse changes on any measure for those who self-harmed at the same rate. For those who had not harmed or did so at a reduced rate, there was a 7.48 postuse reduction on mean MFQ (95% CI 1.94-13.03; t_{24}=2.78; *P*=.01), 16 mean reduction on the RCADS (95% CI 7.63-24.37; t_{24}=3.95, *P*=.001), and 1.0 mean reduction on the emotional subscale of the SDQ (95% CI 0.27-1.73, t_{24}=2.81, *P*=.01).

A total of 18 parents of younger children completed baseline measures. Of these, 13 completed postuse questionnaires with paired data being available for 10-13 parents according to the specific items analyzed. There were no significant differences on the parent-completed RCADS. However, on the SDQ, there was a statistically significant postuse mean difference of 0.92 on the emotional subscale (95% CI 0.083-1.75; t_{11}=2.42; *P*=.03), 2.61 on the peer relationship subscale (95% CI 0.334-4.90; t_{12}=2.50; *P*=.03), and 4.08 on the total score (95% CI 2.16-6.01; t_{11}=4.67; *P*=.001).

Numbers were too small to compare parental reported changes in symptoms for young people who reported changes in self-harm with those who were self-harming at the same rate.

Acceptability

Finally, postuse ratings on a 10-point scale (higher score=more positive endorsement) were high for ease of use (8.9, SD=1.2) and whether they would recommend BlueIce (8.6, SD=1.6) and slightly lower for helpfulness (6.6, SD=2.2). At the end of the study, 29 of the 33 interviewed (88%) wanted to keep the app. Of those who did not want it, 1 no longer felt she needed it, 2 were not ready to stop self-harming, and 1 felt it was too much of a chore to use.

Self-Harm

Self-report changes in self-harm are summarized in Table 5.

All of those who reported not self-harming in the 4 weeks before baseline assessment maintained their status and had not self-harmed over the course of the 12-week trial. Of those 26 who had self-harmed at baseline, 4 (15%) had completely stopped, with a further 15 (58%) reporting less frequent acts of self-harm at follow-up. There was a small group of young people (7, 27%) who reported no reductions in their self-harming behavior over the 12-week trial.

We calculated the number of self-harm acts prevented in 2 ways. First, we obtained data on baseline rates of self-harm for the 4 young people who had stopped self-harming at follow-up. We used this to estimate the number of self-harm events that were prevented if they continued to self-harm at the same rate, during the 12-week study. Second, for the 15 who continued to self-harm but at a reduced rate, we quantified the change to determine how many episodes BlueIce had prevented. Of the 19 young people who reported reduced self-harm, 10 had a reduction of 1-5 incidents, 2 a reduction of 6-10 incidents, and 7 a reduction of 11 or more incidents. These calculations suggest that a total of 308 incidents of self-harm were prevented during the course of this study.

Finally, feedback from young people at the end of the study revealed that no one had called the emergency numbers provided on BlueIce. Similarly, no clinician withdrew any young person from the study due to an escalation of risk or the emergence of suicidal planning or a suicide attempt.

Discussion

Principal Findings

Our study group comprised young people with chronic mental health problems that were significantly impairing most parts of their everyday life. They reported high levels of symptoms of depression and anxiety, with over two-thirds self-harming at least once in the 4 weeks before joining the study. It is therefore encouraging that a majority of this group found BlueIce to be

acceptable and were keen to use it after familiarization and elected to keep it at the end of the study. Similarly, BlueIce proved acceptable to all core professionals groups working in CAMHS. Child and adolescent psychiatrists, clinical psychologist, family therapists, child psychotherapists, nurses, and occupational therapists all referred young people to the project. Statistically significant reductions in symptoms of depression and anxiety over the course of the 12-week trial are also encouraging.

Codesign

The acceptability of BlueIce was undoubtedly enhanced by the involvement of young people and clinicians during the design and development phase. The app was coproduced with young people who advised on all aspects including design, layout, flow, and content. Similarly, workshops with clinicians ensured that the mood-lifting techniques reflected both the recommended interventions of the National Institute for Health and Care Excellence and clinical practice. Our experience supports the recommendation of a recent review that any new therapeutic intervention should be developed in collaboration with patients to ensure that it meets their needs [11].

Adverse Effects

mHealth researchers have reported that apps can pose risks to patient safety and that actions should be taken to mitigate these risks [26,27]. Given the nature of self-harm and associated comorbid difficulties, it was necessary to ensure BlueIce worked as intended and did not make young people's difficulties worse. There were no adverse events during the course of the study. No young person used the emergency contact numbers, none were withdrawn by their clinician following increased risk, and young people did not feel that BlueIce would increase their self-harm. Young people were more cautious when answering about whether BlueIce would help them to stop self-harming. Although only a few completely stopped, almost three-quarters reported that they had stopped or were self-harming at a reduced rate after using BlueIce. These young people identified a total of 308 potential episodes of self-harm that were prevented over a 12-week period. Although the majority of self-harm does not result in emergency department attendance, it is likely that some of these episodes would have resulted in contact with primary or secondary care services. Furthermore, the use of BlueIce may also reduce the number of face-to-face contacts with mental health services. This needs to be examined in future studies as the potential cost savings from this low-cost intervention could be considerable.

Who Did Not Find BlueIce Helpful

There were a small minority of young people who did not find BlueIce helpful. The 3 who declined to use BlueIce after familiarization reported that they did not feel ready to stop self-harming. Similarly, feedback from those who used BlueIce suggested that there were times when their distress was too overwhelming to use the app. Our findings suggest that the preparedness of young people to stop self-harming needs careful assessment. Providing BlueIce for young people who are not motivated or ambivalent to change will not be effective. Second, young people need to be encouraged to use BlueIce at an earlier stage in the distress escalation cycle. By earlier use, they may be able to prevent the distress buildup and self-harm cycle. However, we are mindful that BlueIce is a smartphone app and that the limitations of digital support in preventing all self-harm need to be recognized.

Limitations

Our study is exploratory and as such has a number of limitations. First, we report a self-selected case series of young people who actively elected to trial BlueIce. Our group may, therefore, have been more motivated and prepared to address their self-harm. Second, this was an open uncontrolled study, and we did not have a comparison group. We do not know whether the improvements we report are better or worse than usual care. Third, BlueIce was provided in addition to usual face-to-face intervention. Consequently, we do not know whether the improvements on symptoms of depression, anxiety, and reduction in self-harm we noted were due to BlueIce or their face-to-face intervention. A randomized trial comparing treatment as usual with and without BlueIce is required to quantify the benefits of adding BlueIce to usual care.

Finally, we assessed self-harm through retrospective self-reports, which may be subject to recall bias. Although this limitation is recognized, young people were clearly able to identify changes in their self-harm during the interview, and they directly attributed these to the use of BlueIce rather than to the usual care they had received. Subsequent studies should explore the use of prospective self-harm diaries and the exploration of hospital records of A&E attendances following self-harm.

Finally, we are reporting a small exploratory study that is not sufficiently powered to explore the effects on psychological functioning or self-harm over time. Although our results are promising, we need to be cautious about the strength of any conclusions that can be drawn from our findings. This needs to be addressed in a subsequent, suitably powered, randomized controlled trial.

Conclusions

In spite of these limitations, our results are encouraging and suggest that a smartphone app, used alongside a traditional CAMHS face-to-face intervention, can help young people manage their emotional distress and urges to self-harm. BlueIce proved highly acceptable to young people and clinicians. It was safe with young people reporting a number of improvements in their mood and reductions in self-harm after use. Further research is required to determine the additional benefits of adding BlueIce to face-to-face care and in quantifying potential cost savings that may result from fewer emergency department attendances. However, in the meantime, as an adjunct to usual face-to-face care, BlueIce appears an acceptable and readily accessible way of empowering young people to manage their distress.

Acknowledgments

This project was supported by awards from the Health Foundation (2143 Oxford Health NHS FT) and NHSE Digital. The funders had no role in the study design or in the writing of this manuscript.

Authors' Contributions

PS is the grant holder and principal investigator for the project. PS conceptualized the study design and drafted the manuscript. RG and JP are the researchers involved in the study and collected the research data. All authors read, contributed to, and approved the final manuscript.

Conflicts of Interest

PS designed BlueIce but receives no financial benefit. JP and RG have no competing interests.

References

1. National Collaborating Centre for Mental Health. NICE. 2011. Self-harm in over 8s: long-term management URL: https://www.nice.org.uk/guidance/cg133 [accessed 2018-01-22] [WebCite Cache ID 6so0ErY4a]
2. Hawton K, Saunders KEA, O'Connor RC. Self-harm and suicide in adolescents. Lancet 2012 Jun 23;379(9834):2373-2382. [doi: 10.1016/S0140-6736(12)60322-5] [Medline: 22726518]
3. Kidger J, Heron J, Lewis G, Evans J, Gunnell D. Adolescent self-harm and suicidal thoughts in the ALSPAC cohort: a self-report survey in England. BMC Psychiatry 2012 Jun 27;12:69 [FREE Full text] [doi: 10.1186/1471-244X-12-69] [Medline: 22737989]
4. Hawton K, Rodham K, Evans E, Weatherall R. Deliberate self harm in adolescents: self report survey in schools in England. Br Med J 2002 Nov 23;325(7374):1207-1211 [FREE Full text] [Medline: 12446536]
5. Madge N, Hewitt A, Hawton K, de Wilde EJ, Corcoran P, Fekete S, et al. Deliberate self-harm within an international community sample of young people: comparative findings from the Child & Adolescent Self-harm in Europe (CASE) Study. J Child Psychol Psychiatry 2008 Jun;49(6):667-677. [doi: 10.1111/j.1469-7610.2008.01879.x] [Medline: 18341543]
6. Klonsky ED, Muehlenkamp JJ. Self-injury: a research review for the practitioner. J Clin Psychol 2007 Nov;63(11):1045-1056. [doi: 10.1002/jclp.20412] [Medline: 17932985]
7. Stallard P, Spears M, Montgomery AA, Phillips R, Sayal K. Self-harm in young adolescents (12-16 years): onset and short-term continuation in a community sample. BMC Psychiatry 2013 Dec 02;13:328 [FREE Full text] [doi: 10.1186/1471-244X-13-328] [Medline: 24294921]
8. Campbell D. Theguardian. 2016. NHS figures show 'shocking' rise in self-harm among young URL: https://www.theguardian.com/society/2016/oct/23/nhs-figures-show-shocking-rise-self-harm-young-people [accessed 2018-01-22] [WebCite Cache ID 6tH6jUbna]
9. NSPCC. 2016. Rise in children hospitalised for self-harm as thousands contact Childline URL: https://www.nspcc.org.uk/what-we-do/news-opinion/rise-children-hospitalised-self-harm-thousands-contact-childline/ [accessed 2017-09-06] [WebCite Cache ID 6tH7B97Vm]
10. McMahon EM, Keeley H, Cannon M, Arensman E, Perry IJ, Clarke M, et al. The iceberg of suicide and self-harm in Irish adolescents: a population-based study. Soc Psychiatry Psychiatr Epidemiol 2014 Dec;49(12):1929-1935. [doi: 10.1007/s00127-014-0907-z] [Medline: 24929354]
11. Hawton K, Witt KG, Taylor Salisbury TL, Arensman E, Gunnell D, Townsend E, et al. Interventions for self-harm in children and adolescents. Cochrane Database Syst Rev 2015 Dec 21(12):CD012013. [doi: 10.1002/14651858.CD012013] [Medline: 26688129]
12. Hawton K, Saunders KE, O'Connor RC. Self-harm and suicide in adolescents. Lancet 2012 Jun 23;379(9834):2373-2382. [doi: 10.1016/S0140-6736(12)60322-5] [Medline: 22726518]
13. NHS. Gov.uk. 2015. Future in mind: promoting, protecting and improving our children and young people's mental health and wellbeing URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/414024/Childrens_Mental_Health.pdf [accessed 2018-01-22] [WebCite Cache ID 6tH9tEK0l]
14. Ofcom. 2014. Children and parents: media use and attitudes report 2014 URL: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-parents-oct-14 [accessed 2017-09-06] [WebCite Cache ID 6tHA4H8mN]
15. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
16. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
17. Developer. Developer.nhs.uk. Digital Assessment Questions – Beta URL: https://developer.nhs.uk/digital-tools/daq/ [accessed 2017-09-06] [WebCite Cache ID 6tH7sEKaX]
18. Oxford Health NHS Foundation Trust. Oxfordhealth.nhs.uk. BlueIce app URL: https://www.oxfordhealth.nhs.uk/blueice/ [accessed 2018-01-24] [WebCite Cache ID 6wi17SU9y]

19. Stallard P, Porter J, Grist R. Safety, acceptability, and use of a smartphone app, BlueIce, for young people who self-harm: protocol for an open phase I trial. JMIR Res Protoc 2016 Nov 16;5(4):e217 [FREE Full text] [doi: 10.2196/resprot.6525] [Medline: 27852560]
20. Costello EJ, Angold A. Scales to assess child and adolescent depression: checklists, screens, and nets. J Am Acad Child Adolesc Psychiatry 1988 Nov;27(6):726-737. [doi: 10.1097/00004583-198811000-00011] [Medline: 3058677]
21. Daviss WB, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA. Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. J Child Psychol Psychiatry 2006 Sep;47(9):927-934. [doi: 10.1111/j.1469-7610.2006.01646.x] [Medline: 16930387]
22. Kent L, Vostanis P, Feehan C. Detection of major and minor depression in children and adolescents: evaluation of the Mood and Feelings Questionnaire. J Child Psychol Psychiatry 1997 Jul;38(5):565-573. [Medline: 9255700]
23. Wood A, Kroll L, Moore A, Harrington R. Properties of the Mood and Feelings Questionnaire in adolescent psychiatric outpatients: a research note. J Child Psychol Psychiatry 1995 Feb;36(2):327-334. [Medline: 7759594]
24. Chorpita BF, Moffitt CE, Gray J. Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. Behav Res Ther 2005 Mar;43(3):309-322. [doi: 10.1016/j.brat.2004.02.004] [Medline: 15680928]
25. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 1997 Jul;38(5):581-586. [Medline: 9255702]
26. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
27. Naeem F, Gire N, Xiang S, Yang M, Syed Y, Shokraneh F, et al. Reporting and understanding the safety and adverse effect profile of mobile apps for psychosocial interventions: an update. World J Psychiatry 2016 Jun 22;6(2):187-191 [FREE Full text] [doi: 10.5498/wjp.v6.i2.187] [Medline: 27354959]

Abbreviations

CAMHS: child and adolescent mental health services
CBT: cognitive-behavioral therapy
DBT: dialectical behavioral therapy
MFQ: Mood and Feelings Questionnaire
NHS: National Health Service
RCADS: Revised Child Anxiety and Depression Scale
SDQ: Strengths and Difficulties Questionnaire

Edited by G Eysenbach; submitted 08.09.17; peer-reviewed by S Merry, S Hetrick, G Pramana, H Miller; comments to author 25.10.17; revised version received 07.11.17; accepted 17.11.17; published 30.01.18.

Please cite as:
Stallard P, Porter J, Grist R
A Smartphone App (BlueIce) for Young People Who Self-Harm: Open Phase 1 Pre-Post Trial
JMIR Mhealth Uhealth 2018;6(1):e32
URL: http://mhealth.jmir.org/2018/1/e32/
doi:10.2196/mhealth.8917
PMID:29382632

Original Paper

A Mobile App to Screen for Neurocognitive Impairment: Preliminary Validation of NeuroScreen Among HIV-Infected South African Adults

Reuben N Robbins[1*], PhD; Hetta Gouse[2*], PhD; Henry G Brown[3*], BSc (Hons); Andries Ehlers[3*], BTech; Travis M Scott[4*], MA; Cheng-Shiun Leu[1*], PhD; Robert H Remien[1*], PhD; Claude A Mellins[1*], PhD; John A Joska[2*], MBChB, PhD

[1]HIV Center for Clinical and Behavioral Studies, Department of Psychiatry, New York State Psychiatric Institute & Columbia University, New York, NY, United States

[2]HIV Mental Health Research Unit, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa

[3]Envisage IT, Cape Town, South Africa

[4]Department of Psychology, Fordham University, Bronx, NY, United States

[*]all authors contributed equally

Corresponding Author:
Reuben N Robbins, PhD
HIV Center for Clinical and Behavioral Studies
Department of Psychiatry
New York State Psychiatric Institute & Columbia University
1051 Riverside Drive, Unit 15
New York, NY, 10032
United States
Phone: 1 6467746968
Fax: 1 6467746955
Email: rnr2110@cumc.columbia.edu

Abstract

Background: Neurocognitive impairment (NCI) is one of the most common complications of HIV infection, and has serious medical and functional consequences. South Africa has 7 million people living with HIV (PLHIV) with up to three-quarters of antiretroviral therapy (ART)-naïve individuals having NCI. South Africa's health system struggles to meet the care needs of its millions of PLHIV; screening for NCI is typically neglected due to limited clinical staff trained to administer, score, and interpret neuropsychological tests, as well as long test batteries and limited screening tools for South African populations. Without accurate, clinically useful, and relatively brief NCI screening tests that can be administered by all levels of clinical staff, critical opportunities to provide psychoeducation, behavioral planning, additional ART adherence support, and adjuvant therapies for NCI (when they become available) are missed. To address these challenges and gap in care, we developed an mHealth app screening tool, NeuroScreen, to detect NCI that can be administered by all levels of clinical staff, including lay health workers.

Objective: The purpose of this study was to examine sensitivity and specificity of an adapted version of NeuroScreen to detect NCI (as determined by a gold standard neuropsychological test battery administered by a trained research psychometrist) among HIV-infected South Africans when administered by a lay health worker.

Methods: A total of 102 HIV-infected black South African adults who had initiated ART at least 12 months prior were administered NeuroScreen and a gold standard neuropsychological test battery in the participants' choice of language (ie, English or isiXhosa). Three composite z scores were calculated for NeuroScreen: (1) sum of all individual test scores, (2) sum of all individual test scores and error scores from four tests, and (3) sum of four tests (abbreviated version). Global deficit scores were calculated for the gold standard battery where a score of 0.5 or greater indicated the presence of NCI.

Results: The mean age of participants was 33.31 (SD 7.46) years, most (59.8%, 61/102) had at least 12 years of education, and 81.4% (83/102) of the sample was female. Gold standard test battery results indicated that 26.5% (27/102) of the sample had NCI. Sensitivity and specificity of age-, education-, and sex-adjusted NeuroScreen scores were 81.48% and 74.67% for composite score 1, 81.48% and 81.33% for composite score 2, and 92.59% and 70.67% for composite score 3, respectively.

Conclusions: NeuroScreen, a highly automated, easy-to-use, tablet-based screening test to detect NCI among English- and isiXhosa-speaking South African HIV patients demonstrated robust sensitivity and specificity to detect NCI when administered by lay health workers. NeuroScreen could help make screening for NCI more feasible. However, additional research is needed with larger samples and normative test performance data are needed.

(JMIR Mhealth Uhealth 2018;6(1):e5) doi:10.2196/mhealth.9148

KEYWORDS

HIV; neurocognitive; impairment; lay health workers; resource-limited settings; South Africa; tablet; app; neuropsychology

Introduction

Neurocognitive impairment (NCI) is one of the most common sequelae and comorbid conditions of HIV infection and has significant medical, functional, and public health consequences. Prevalence estimates of NCI among South Africa's 7 million people living with HIV (PLHIV) [1] range from 23% to 76% depending on whether the sample is antiretroviral therapy (ART) experienced or naïve, and do not differ between men and women [2,3]. These rates are similar to estimates in high-income countries [4-8]. Neurocognitive impairment in HIV, known as HIV-associated neurocognitive disorder (HAND) when HIV is determined to be the etiology, typically causes impairments in mental processing speed, learning, memory, attention and concentration, executive functions, and motor speed [6,9,10]. It ranges in severity from asymptomatic and mild forms to a severe dementia-type form [11]. Having even asymptomatic or mild NCI has been associated with increased risk for developing more severe impairment and mortality [12-16]. Neurocognitive impairment has also been associated with worse ART adherence (jeopardizing positive health outcomes), employment difficulties, impaired instrumental activities of daily living (eg, planning, driving, finance management), worse overall quality of life, and the need for more social services [12-23]. In addition, NCIs are also associated with poor decision making and greater HIV transmission risk behaviors (eg, unprotected sex) [17-19].

Screening for NCI in HIV is essential to good comprehensive care and treatment strategies [8,20,21]. As a first step in the clinical decision-making process, screening can enable providers to determine who is most likely to have NCI, detect early signs of NCI, allocate limited resources more effectively (eg, comprehensive neuropsychological testing, neurologic services, and social worker intervention), determine when to initiate and adjust ART regimens, track and monitor neurocognitive function, and educate patients about the impact of NCI and ways to minimize it—all which can improve health outcomes [22,23]. The impact of NCI on ART adherence can be minimized through comprehensive behavioral planning that incorporates reminder strategies and social support. Detecting NCI in highly infectious PLHIV (ie, those with detectable viral load) may help in tailoring transmission prevention strategies. Furthermore, if and when adjuvant pharmacotherapies or behavioral interventions become available for NCI in HIV, screening for NCI will be the first step to linking patients with these treatments. However, screening for NCI among PLHIV is not in widespread practice, and faces unique challenges in low- and middle-income countries and resource-limited settings. In South Africa, there are few locally developed screening tests for NCI in the local languages of those most affected by HIV, and a lack of expert personnel to administer them [13,24-28], missing critical opportunities to detect NCI and intervene for the millions of PLHIV there.

A variety of neurocognitive screening tests have been developed and evaluated to detect HAND, including computer-based and paper-and-pencil tests. Recent reviews of the most commonly used screening tests for HAND, mostly from high-income countries, indicated a wide range of sensitivity (<55%-90%) and specificity (70%-90%) depending on the test, the specific population, and type of NCI detected (ie, mild or HIV-associated dementia) [29]. Studies examining NCI screening tests used specifically with South African PLHIV found similarly wide variation in sensitivity and specificity depending on the test or combination of tests used [30,31]. None of these studies evaluated the performance of the screening tests when administered by nonspecialists or highly trained personnel. A screening test that can be administered by all levels of clinical staff is essential for scale-up in low- to middle-income countries and other resource-limited settings.

In South Africa, like many other low- and middle-income countries and resource-limited settings, lay health workers are utilized in clinical settings to provide essential HIV-related services (eg, AIDS education, HIV testing and counseling, ART adherence counseling, and preliminary management of mild mental disorders) that may not be otherwise available due to provider shortages [28,32-34]. Through task shifting, lay health workers work alongside nurses, physicians, pharmacists, and social workers. Lay health workers are ideally situated to provide a number of services to PLHIV, but certain routine tasks are still too complex, such as neurocognitive testing and screening [27].

To address this gap in HIV care, we developed NeuroScreen [35], an mHealth app for tablets designed to be used by all levels of health care personnel, including lay health workers, to screen for NCI. This novel software app for the Android operating system uses the touchscreen capabilities of tablet devices to highly automate neuropsychological testing. The neurocognitive screening test battery is embedded in a graphical user interface that automates test administration and allows for easy data management and reporting. All tests in NeuroScreen are automatically timed and scored—no hand scoring, score converting, or simultaneous and synchronized use of stopwatches is required. Tests that would normally require a pen or pencil use the touchscreen instead. To ensure that each administration is consistent, administrators are forced to sequence through all the standardized instructions. Furthermore, eight of the ten tests have audio-visual instructions, which are

useful for low-literacy populations. Because NeuroScreen runs on a tablet and does not require an Internet connection, it is ultraportable and allows screenings to be administered in almost any location, such as remote or rural clinics or fast-paced and busy urban clinics requiring flexible use of examination rooms. Although computerized neuropsychological testing is becoming more common (but not so in resource-limited settings), mobile operating systems and devices are only just starting to be more commonly used as test delivery platforms.

In our previous work [35], we pilot tested a large-format mobile phone version of NeuroScreen in a small sample (N=44) of older (mean age 53.4 years) HIV-positive adults in the United States, most of whom (75%) had NCI as defined by a gold standard neuropsychological test battery. We also assessed acceptability of NeuroScreen among participants and 10 HIV providers. Evidence for construct validity of individual NeuroScreen tests as compared to the gold standard neuropsychological test battery was established. NeuroScreen also showed high sensitivity and moderate specificity to detect NCI at 93.94% and 63.64%, respectively. Both participants and providers indicated high acceptability for NeuroScreen. Providers remarked that NeuroScreen has potential to detect other disease-related NCIs.

The purpose of this study was to evaluate the ability of the lay health worker-administered NeuroScreen to detect NCI, as defined by a gold standard neuropsychological test battery. We examined the sensitivity and specificity, as well as the negative predictive value (NPV) and positive predictive value (PPV) of a South African-adapted NeuroScreen (for use with tablets and with the isiXhosa language) to detect NCI.

Methods

This cross-sectional study adapted NeuroScreen from a large mobile phone screen format to 7-inch tablet format. Specifically, the Google Nexus 7 tablet was used for this study. Both the user interface and tests were updated to take advantage of the larger screen. NeuroScreen was also adapted for isiXhosa-speaking South Africans (the predominant Bantu language spoken in the Western Cape region of South Africa, where this study was conducted). Language translations underwent forward and backward translation, as well as vetting by the bilingual (English and isiXhosa) study staff. IsiXhosa translations were optimized to follow colloquial language conventions because it is the more predominantly used and understood form of isiXhosa in the communities where the study took place. Audio files were recorded with a native isiXhosa speaker for the audio-visual instructions. After all updates were made, the software engineers conducted quality assurance testing.

Sample and Recruitment

A total of 102 HIV-positive black South African adults, aged 18 to 56 years, were recruited following their participation in a larger randomized controlled trial (RCT) of a multimedia, laptop-based, lay health worker-delivered ART-readiness intervention for ART initiators (known as Masivukeni or "Let's Wake Up") conducted in Cape Town, South Africa (see [36-38]). Inclusion criteria were (1) HIV-positive, (2) age 18 years or older, (3) willing to be contacted by study staff for participation in other research studies, (4) isiXhosa or English speaking, (5) capable of consent, (6) willing to complete 2 to 3 hours of neuropsychological testing, and (7) willing to allow access to medical records by study staff. Exclusion criteria were (1) not meeting one of the preceding criteria, and (2) presence of a current psychotic disorder, significant current suicidal ideation, and severe cognitive impairment precluding ability to give informed consent or participate based on clinical judgment of providers. All participants enrolled in the larger RCT were eligible to participate in this study. Ethics approval was obtained by the Human Research Ethics Committee at the University of Cape Town and the New York State Psychiatric Institute Institutional Review Board.

Procedure

When the RCT participants completed their final study visit, the study nurse informed them about an additional study and requirements for participation. If a participant was interested and eligible, written informed consent was obtained. Then a trained lay health worker (two at each of the study clinics) administered NeuroScreen. A trained neuropsychology technician (one at each of the two study clinics) then administered the gold standard neuropsychological battery (approximately 3 hours). For participants who were unable to complete all study requirements on that visit, the neuropsychology technician scheduled another visit within 7 days to complete the gold standard neuropsychological battery. Neuropsychology technicians were blind to NeuroScreen results. Participants received the equivalent of US $40 for completion of all study procedures.

Measures

Demographic and Neuromedical Data

Demographic and medical history data were available from the larger RCT. The psychometrist conducted an additional neuromedical questionnaire prior to administering the neuropsychological battery.

NeuroScreen

NeuroScreen [35] is comprised of 10 brief neuropsychological tests to briefly assess individuals across six neuropsychological domains most affected by HIV: verbal learning (two trials, five words) and memory (5-minute delayed recall); processing speed via two trail making test sequencing paradigms, two visual discrimination tasks, and a number input task; attention/concentration via a number span forward and backward task; executive functioning via an alternating trail making test sequencing paradigm; and motor functioning via a finger-tapping task (see Multimedia Appendix 1 for a complete description of each test). The tests are embedded in a graphical user interface that allows the administrator to enter patient data, administer tests, generate instant raw results, and save raw results to a secure and password-protected website and/or an internal storage card. After entering patient data, the administrator is required to read standardized test instructions or play videos that provide audio-visual instructions, is prompted at appropriate points to offer practice trials on selected tests, and is prompted to move on to the next test, thus sequencing through all the tests.

NeuroScreen can be administered in English or isiXhosa. Once all data were transferred to the principal investigator's secure and encrypted hard drive, all data were wiped from the device.

Total raw scores (eg, completion times, total correct) and error scores (eg, sequencing errors in the trail making tests and incorrect number inputs in the number input test) are systematically captured by the app. For this study, all raw scores were converted to z scores using the entire sample and timed tests were reverse scored so that higher scores indicated better performance. Three composite z scores were calculated: (1) sum of all individual test scores, (2) sum of all individual test scores and total errors from the trail making and number speed tests, and (3) sum of four tests (visual discrimination 1 and 2, trail making 1, and number span total).

Neuropsychological Evaluation

This paper-and-pencil neuropsychological battery assessed individuals across all neurocognitive domains using tests that are particularly sensitive to HIV-related NCI: learning and memory (Hopkins Verbal Learning Test-Revised [HVLT-R] [39] and the Brief Visuospatial Memory Test-Revised [BVMT-R] [40]); executive functioning (Color Trails Test 2 [41] and the Wisconsin Card Sorting Test [WCST] [42]); attention/concentration (Wechsler Memory Scales, Third Edition [WMS-III] Spatial Span [43] and Wechsler Adult Intelligence Scales, Third Edition [WAIS-III] Digit Span [44]); processing speed (Trail Making Test, Part A [45], Color Trails Test 1 [41], and WAIS-III Digit Symbol Coding and Symbol Search [44]); language (semantic fluency of animals, fruits, and vegetables [46]), and motor functioning (successive finger tapping and grooved pegboard [47]). This particular paper-and-pencil battery has been used in many HIV studies in the United States [7,26,48] and has been adapted for use with isiXhosa speakers in South Africa [2,3]. Data produced by this battery include (1) raw test scores, (2) T scores, (3) individual test deficit scores, and (5) global deficit scores (GDS).

A GDS was calculated for each participant and used to identify those with and without NCI. The GDS summarizes neuropsychological battery results by converting individual test T scores to a deficit score from 0 (no impairment) to 5 (severe impairment) [49]. Deficit scores are averaged across all tests to create the GDS. The GDS considers number and severity of impairments, assigning less weight to unimpaired performance and overcoming the disadvantage of averaging absolute performance, which gives equal weight to unimpaired and impaired scores [50]. The GDS method detects mild NCI (GDS≥0.5) across varying impairment patterns in different neurocognitive domains [49,51]. We set the threshold of GDS≥0.5 as indicative of NCI.

Statistical Analysis

Univariate analyses were conducted to examine participant characteristics. To evaluate the ability of the lay health worker-administered NeuroScreen to detect NCI and evaluate its sensitivity and specificity, a logistic regression and receiver operator characteristic curve were used. First, the logistic regression model was used to establish a prediction model for NCI using the NeuroScreen score and age, gender, and education in the form of: $\log[P(X)/\{1-P(X)\}] = \beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4$, where $P(X)$ is the probability of having NCI given $X=(X_1,\ldots, X_4)$=(patient's age, sex, years of education, and NeuroScreen total score). After obtaining the estimated regression coefficients, the predicted probability of having NCI was calculated. After calculating the predicted probability for each patient, a receiver operating characteristic (ROC) analysis using the predicted probability and the gold standard NCI determination was calculated to evaluate the area under the curve (AUC) and to generate an optimal age-, education-, and sex-adjusted NeuroScreen score cutoff point to distinguish participants with NCI. The cutoff point was maximized by using the Youden index, which is equal to the sensitivity+specificity–1. Both PPV and NPV were computed for the optimized NeuroScreen cutoff score. All analyses were conducted using IBM SPSS version 23 (IBM Corp, Armonk, NY, USA).

Results

Sample Characteristics

As Table 1 shows, the sample was predominantly female with a mean age of 33.31 (SD 7.46) years and most (59.8%, 61/102) had at least 12 years of education. Although 30 participants reported being held back for at least one grade during their schooling, only nine participants reported receiving special classes to assist them with learning difficulties (the South African education system does not routinely assess for and diagnose learning disorders). Four participants reported having had a loss of consciousness greater than 15 minutes (for head injury and other medical issues). One participant reported having a stroke that resulted in numbness in the right arm, three participants reported having had one seizure, and one participant reported being diagnosed with epilepsy at age 10. Four participants reported taking medication for high blood pressure.

NeuroScreen Performance

Table 2 displays raw mean scores for all NeuroScreen tests, as well as the three composite total *z* scores. On average, participants were able to learn 8.45 (SD 1.51) words across two learning trials (same five words per trial) and recall 3.39 (SD 1.20) words after a 5-minute delay. The mean total number span backward and forward score was 5.87 (SD 1.29) with a maximum possible score of 17. The mean total correct responses on Visual Discrimination 1 was 11.28 (maximum=61) and 25.97 on Visual Discrimination 2 (maximum=150). The mean completion time on Number Input Speed was 45.02 (SD 18.08) seconds. Mean completion times for Trails Test 1 was 18.93 (SD 18.17) seconds, 31.95 (SD 21.21) seconds for Trails Test 2, and 18.81 (SD 16.32) seconds for Trails Test 3. The mean total finger taps for both the dominant and nondominant hand was 454.73 (SD 60.03) taps across five trials.

Gold Standard HIV Neuropsychological Battery Performance

Results from the full neuropsychological battery (Table 3) indicated that the sample had a mean global T score of 48.01 (SD 4.79). Table 3 also displays performance (adjusted T scores) across the individual tests. The mean GDS was 0.36 (SD 0.40)

and 26.5% (27/102) had NCI using a GDS score of 0.5 or greater to indicate impairment.

Sensitivity and Specificity

NeuroScreen Total Score 1 (Sum of All Tests)

Using the logistic model with the first NeuroScreen total score adjusted for age, education, and sex to predict the gold standard NCI in the ROC analysis, the AUC was 0.86 (95% CI 0.78-0.94; see Figure 1). The Youden index NeuroScreen predicted NCI cut-score of 0.21 maximized sensitivity at 81.48% (95% CI 61.92%-93.70%) and specificity at 74.67% (95% CI 63.30%-84.01%). The PPV was 53.66% and the NPV was 91.80%. Using this cut-score yielded 19 false positives and 5 false negatives. The mean completion time for all the tests was 23.88 (SD 6.21) minutes.

NeuroScreen Total Score 2 (Sum of All Tests and Available Error Scores)

An ROC analysis using the logistic model with the second NeuroScreen total score to predict NCI had an AUC of 0.86 (95% CI 0.78-0.94; see Figure 2). The Youden index maximal sensitivity was 81.48% (95% CI 61.92%-93.70%) and specificity was 81.33% (95% CI 70.67%-89.40%). The PPV was 61.11% and the NPV was 92.42%. Using this cut-score yielded 14 false positives and 5 false negatives. The mean completion time was the same as the preceding.

Table 1. Sample characteristics (N=102).

Characteristic	Participants	Min	Max
Age (years), mean (SD[a])	33.31 (7.46)	19	56
Gender (female), n (%)	83 (81)	—	—
Education (years completed), mean (SD)	11.25 (1.99)	3	14
Traumatic brain injury with loss of consciousness >15 minutes, n (%)	4 (4)	—	—
Likely learning difficulty, n (%)	9 (9)	—	—
Most recent CD4 cell count, mean (SD)[b]	501.31 (287.41)	47	1654
Most recent viral load undetectable, n (%)[c]	81 (91)	—	—

[a]SD: standard deviation.

[b]CD4: cluster of differentiation 4. CD4 cell count available for 88 participants.

[c]Viral load data available for 81 participants.

Table 2. NeuroScreen performance (raw).

Test	Mean (SD[a])	Min	Max
Finger tapping total (both hands)	454.73 (60.03)	190.00	581.00
Visual discrimination 1 total correct	11.28 (3.67)	4.00	19.00
Visual discrimination 2 total correct	25.97 (7.97)	1.00	48.00
Number span total (forward and backward)	5.87 (1.29)	3.50	9.50
Verbal learning total correct	8.45 (1.51)	3.00	10.00
Delayed verbal recall total correct	3.39 (1.20)	0.00	5.00
Trail making 1 completion time (seconds)[b]	–18.93 (18.17)	–120.00	–6.06
Trail making 2 completion time (seconds)[b]	–31.95 (21.21)	–120.00	–11.17
Trail making 3 completion time (seconds)[b]	–18.81 (16.32)	–120.00	–5.05
Number speed completion time (seconds)[b]	–45.02 (18.08)	–117.75	–23.60
Full battery completion time (minutes)	23.88 (6.21)	9.00	52.00

[a]SD: standard deviation.

[b]Indicates reverse scored (slower time=worse performance).

Table 3. Gold standard neuropsychological test battery performance (adjusted).

Test	Mean (SD[a])	Min	Max
Global T	48.01 (4.79)	34.00	57.63
Motor functioning tests			
Successive finger tapping (dominant hand)	46.51 (11.29)	–1.31	64.05
Successive finger tapping (nondominant hand)	44.98 (13.45)	–43.87	61.66
Grooved pegboard (dominant hand)	47.41 (7.96)	12.67	60.05
Grooved pegboard (nondominant hand)	49.26 (3.09)	34.09	53.27
Hopkins Verbal Learning Test-Revised			
Total trials 1-3	45.97 (8.33)	24.34	62.67
Delay recall total	44.74 (9.86)	22.64	67.13
Brief Visuospatial Memory Test-Revised			
Total Trials 1-3	48.40 (9.80)	28.09	73.75
Delay Total Recall	49.58 (11.48)	29.03	72.20
Wechsler Adult Intelligence Scales, Third Edition (WAIS-III)			
Digit symbol coding total	46.73 (10.08)	25.45	77.15
Symbol search total	47.24 (8.95)	27.19	64.76
Wechsler Memory Scales, Third Edition spatial span total	50.47 (9.65)	30.06	73.85
Processing speed tests			
Trail making test, part A	43.82 (11.99)	–7.48	72.04
Color trails test 1	46.72 (9.98)	12.22	64.02
Color trails test 2	48.32 (9.73)	14.83	66.80
WAIS-III Digit span total	49.33 (1.21)	46.70	52.09
Wisconsin Card Sorting Test			
Perseverative errors	50.74 (12.89)	3.52	63.76
Trials to first sort	48.10 (11.32)	29.66	57.98
Failures to maintain set	50.59 (4.79)	31.88	53.94
Language tests			
Animal fluency total	49.41 (8.70)	27.72	70.48
Fruit and vegetable fluency total	51.83 (8.75)	33.26	72.21

[a]SD: standard deviation.

Figure 1. Receiver operating characteristic curve for NeuroScreen total score 1.

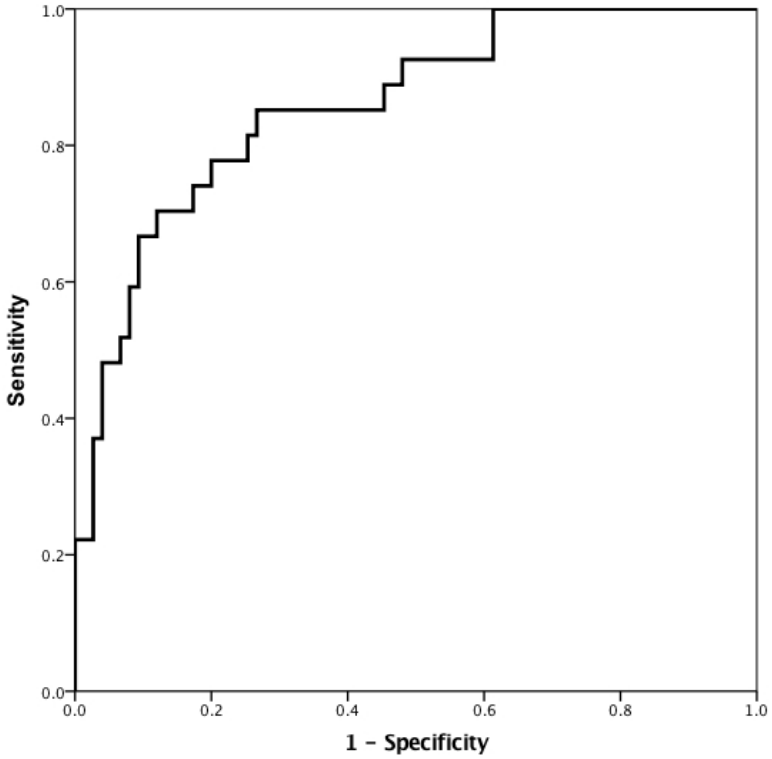


Figure 2. Receiver operating characteristic curve for NeuroScreen total score 2.

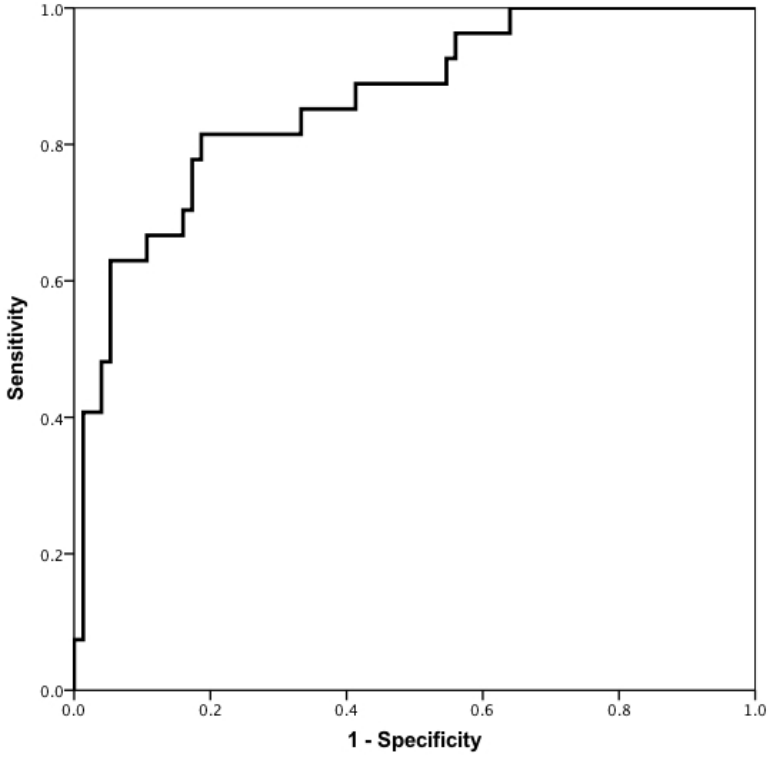
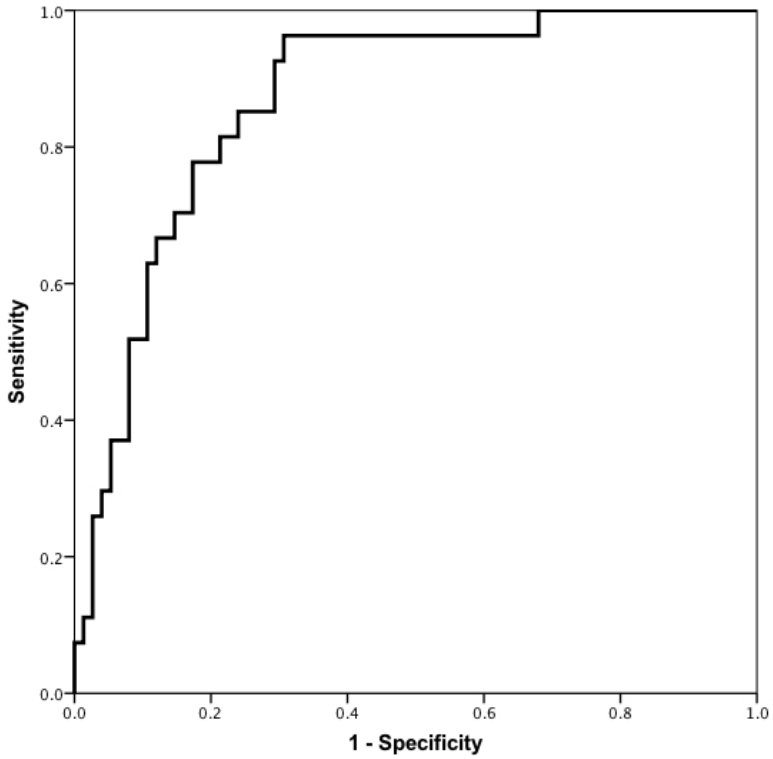


Figure 3. Receiver operating characteristic curve for NeuroScreen total score 3 (abbreviated version).



NeuroScreen Total Score 3 (Sum of Four Tests)

Using the logistic model with the third NeuroScreen total score with age, education, and sex to predict the gold standard NCI in the ROC analysis, the AUC was 0.87 (95% CI 0.80-0.94; see Figure 3). The Youden index NeuroScreen predicted NCI cut-score of 0.18 maximized sensitivity at 92.59% (95% CI 75.71%-99.09%) and specificity at 70.67% (95% CI 59.02%-80.62%). The PPV was 53.19% and the NPV was 96.36%. Using this cut-score yielded 22 false positives and 2 false negatives. The estimated completion time of these four tests was approximately 12 minutes.

Discussion

In a sample of 102 HIV-positive predominantly female South Africans on ART for at least 1 year, gold standard neuropsychological test results indicated 26% of the sample had at least mild NCI. This rate of NCI is consistent with other research from South Africa [3], but lower than NCI rates found in the United States [48] among ART-experienced PLHIV. NeuroScreen—when administered by a lay health worker—had robust test characteristics to detect gold standard-defined NCI ranging from 81% to 93% sensitivity and 71% to 81% specificity, depending on the combination of NeuroScreen test scores used. Our study provides preliminary evidence for the validity of NeuroScreen to detect NCI among English- and isXhosa-speaking South African adults living with HIV. NeuroScreen shows promise as an easy-to-use, brief NCI screening test that can be administered by lay health workers, and likely most levels of health care staff.

Deciding which NeuroScreen total score to use must be weighed by each provider and/or health care system. Given estimated rates of NCI that exist among PLHIV in South Africa (23%-76% [2,3]), the extra burden screening for NCI places on an already overburdened and resource-limited health care system must be weighed against the consequences misdetection might have on patients and the health care system. Ideally, anyone who screens positive for NCI would be referred for confirmatory testing with a comprehensive neuropsychological assessment. However, doing so is simply not feasible in many settings, such as South Africa. For resource-limited settings, such as South Africa, it may make sense to use the shortest version of NeuroScreen (12 minutes to administer with 92.59% sensitivity and 70.67% specificity) even though the rate of false positives would be higher than using the full NeuroScreen with error scores (25-minute administration time with 81.48% sensitivity and 81.33% specificity). The consequences of a positive screen (eg, extra treatment planning considerations and additional ART adherence support) might outweigh the consequences of missing someone who truly has NCI (eg, poor adherence/health outcomes).

Although the full version with error scores is approximately twice as long to administer than the abbreviated version, its specificity is much higher. As a potential substitute for a full neuropsychological assessment (which is not feasible and highly unlikely to occur in this setting), 25 minutes of a lay health worker's time is much less resource intensive than 3 to 5 hours of a neuropsychologist's, psychologist's, and/or neuropsychology technician's time to administer a neuropsychological test battery and score it. Although we do not believe a short battery of tests such as NeuroScreen should replace gold standard neuropsychological assessments, it may help provide clinics in resource-limited settings that do not have access to the gold standard measures and procedures with a viable alternative. Having a screening test that can be

administered by any staff, lay health worker included, can help clinics identify those PLHIV at highest risk for having NCI and make important treatment recommendations and referrals to address it.

Compared to other computerized and paper-and-pencil screening tests for NCI in HIV, NeuroScreen performed similarly to and better than some tests (see [29]). Compared to screening tests specifically evaluated in South Africa [30,31], NeuroScreen also yielded robust sensitivity and specificity across all three total scores. Moreover, these robust performance characteristics were achieved by NeuroScreen when administered by a lay health worker. NeuroScreen, to our knowledge, is the only computerized, mHealth neurocognitive screening test tablet app for NCI detection in South Africa for English and isiXhosa speakers. Furthermore, it is the only such test designed to be administered by all levels of clinical staff with minimal training and supervision.

It is important to note that there was discrepancy between NeuroScreen and the gold standard neuropsychological battery in defining who had NCI. Overall, NeuroScreen indicated that more participants had NCI than did the gold standard battery. The discrepancy could be explained in part by NeuroScreen's tests being more difficult than the gold standard tests (ie, floor effects on NeuroScreen tests) or practice effects. We did not randomize test sequence administration—all participants were administered NeuroScreen first then the gold standard tests. Some participants' performances on the gold standard tests could have benefited from first taking the NeuroScreen tests. If the benefit was big enough, performance on the gold standard tests could have appeared within the normal range (ie, not impaired). Additionally, the discrepancy could also be due to human factors involved in the gold standard battery administration, such as time keeping and recording errors, or subtle biases in test scoring (although our neuropsychology technicians received intensive training and ongoing supervision). There could also be issues regarding language—more participants opted to take NeuroScreen in English (n=27) than the gold standard battery (n=6). Further research is needed to fully understand this discrepancy.

It is important to note this study's limitations. First, we had a small sample of PLHIV in South Africa to evaluate NeuroScreen's sensitivity and specificity that was mostly female. Second, we did not formally assess language fluency, for either English or isiXhosa. Participants were asked for their language preference by the lay health worker. Similarly, the psychometrist discussed with the participant in which language they would like to take the gold standard battery. Third, the NCI detected in this study may or may not be a result of HIV—numerous factors can cause and/or contribute to the development of NCIs, such as low education, head injuries, and other medical factors (many of which were observed in this sample). Fourth, normative performance data have not been established for NeuroScreen among isiXhosa-speaking South Africans, or other South African language groups, making generalization of performance on it inappropriate. Finally, the gold standard battery had numerous neuropsychological tests assessing neurocognitive domains that NeuroScreen did not assess (verbal fluency and perseveration), although the tests in NeuroScreen were specifically chosen to assess those neurocognitive domains most typically affected by HIV. Furthermore, NeuroScreen is not meant to be a substitute for a thorough neuropsychological assessment.

Despite these limitations, we believe NeuroScreen has potential to offer busy health clinics and research studies with a brief, easy-to-use solution to screen for NCI among PLHIV, and among patients with other brain-involving diseases and disorders. With a tool such as NeuroScreen, better referrals, tracking, and integration with electronic medical records could be achieved. However, more research is needed to validate NeuroScreen as a screening tool for NCI. A larger sample, statistically powered to establish internal and external validity indicators, is essential for NeuroScreen scale-up.

Computerized neurocognitive testing and screening are transforming clinical practice for neuropsychological assessments. Mobile technology offers a powerful platform that is ultraportable and can be easy to use. This study provides evidence that our app, NeuroScreen, has clinically useful psychometric properties to detect NCI when administered by lay health workers. Taking advantage of mobile platforms and automating many components of the neurocognitive testing process may help to make testing more accurate, efficient, affordable, and accessible to those who need testing, especially in resource-limited settings.

Acknowledgments

This research was supported by grants from the National Institute of Child Health and Human Development (R21-HD084197; principal investigator: RNR), National Institute of Nursing Research (R21-NR015009; principal investigator: RNR), National Institute of Mental Health (R01-MH09557; principal investigator: RHR), National Institute of Mental Health to the HIV Center for Clinical and Behavioral Studies at New York State Psychiatric Institute and Columbia University (P30-MH43520; principal investigator: RHR), as well as the Columbia University Global Mental Health Scholars Program (director: Kathy Pike, PhD). The authors also wish to acknowledge the hard work of our research staff—Michelle Henry, Veronica Jonah, Anna Dreyer, Yoliswa Mtingeni, Thandeka Mbonambi, Zodwa Mbakaza, Tandiwe Mngxuma, Ziyanda Ncusane, and Zodwa Makuluma—and all our study participants. We would like to thank the City of Cape Town Department of Health for their support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

NeuroScreen test descriptions.

[PDF File (Adobe PDF File), 169KB - mhealth_v6i1e5_app1.pdf]

References

1. UNAIDS. 2017 Oct 10. AIDSInfo URL: http://aidsinfo.unaids.org/ [accessed 2017-10-10] [WebCite Cache ID 6u6ydn7i0]
2. Joska JA, Westgarth-Taylor J, Myer L, Hoare J, Thomas KG, Combrinck M, et al. Characterization of HIV-associated neurocognitive disorders among individuals starting antiretroviral therapy in South Africa. AIDS Behav 2011 Aug;15(6):1197-1203. [doi: 10.1007/s10461-010-9744-6] [Medline: 20614176]
3. Joska JA, Westgarth-Taylor J, Hoare J, Thomas KG, Paul R, Myer L, et al. Neuropsychological outcomes in adults commencing highly active anti-retroviral treatment in South Africa: a prospective study. BMC Infect Dis 2012 Feb 15;12:39 [FREE Full text] [doi: 10.1186/1471-2334-12-39] [Medline: 22335937]
4. Becker JT, Lopez OL, Dew MA, Aizenstein HJ. Prevalence of cognitive disorders differs as a function of age in HIV virus infection. AIDS 2004 Jan 01;18 Suppl 1:S11-S18. [Medline: 15075493]
5. Bonnet F, Amieva H, Marquant F, Bernard C, Bruyand M, Dauchy FA, S CO3 Aquitaine Cohort. Cognitive disorders in HIV-infected patients: are they HIV-related? AIDS 2013 Jan 28;27(3):391-400. [doi: 10.1097/QAD.0b013e32835b1019] [Medline: 23079813]
6. Grant I, Sacktor N, McArthur J. HIV neurocognitive disorders. In: Gendelman HE, Grant I, Everall IP, Fox HS, Gelbard HA, Lipton SA, et al, editors. The Neurology of AIDS. Oxford: Oxford University Press; 2005:357-373.
7. Heaton RK, Grant I, Butters N, White DA, Kirson D, Atkinson JH, et al. The HNRC 500--neuropsychology of HIV infection at different disease stages. HIV Neurobehavioral Research Center. J Int Neuropsychol Soc 1995 May;1(3):231-251. [Medline: 9375218]
8. Simioni S, Cavassini M, Annoni J, Rimbault AA, Bourquin I, Schiffer V, et al. Cognitive dysfunction in HIV patients despite long-standing suppression of viremia. AIDS 2010 Jun 01;24(9):1243-1250. [doi: 10.1097/QAD.0b013e3283354a7b] [Medline: 19996937]
9. Grant I. Neurocognitive disturbances in HIV. Int Rev Psychiatry 2008 Feb;20(1):33-47. [doi: 10.1080/09540260701877894] [Medline: 18240061]
10. Heaton RK, Franklin DR, Ellis RJ, McCutchan JA, Letendre SL, Leblanc S, CHARTER Group, HNRC Group. HIV-associated neurocognitive disorders before and during the era of combination antiretroviral therapy: differences in rates, nature, and predictors. J Neurovirol 2011 Feb;17(1):3-16 [FREE Full text] [doi: 10.1007/s13365-010-0006-1] [Medline: 21174240]
11. Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, et al. Updated research nosology for HIV-associated neurocognitive disorders. Neurology 2007 Oct 30;69(18):1789-1799 [FREE Full text] [doi: 10.1212/01.WNL.0000287431.88658.8b] [Medline: 17914061]
12. Hinkin CH, Castellon SA, Durvasula RS, Hardy DJ, Lam MN, Mason KI, et al. Medication adherence among HIV+ adults: effects of cognitive dysfunction and regimen complexity. Neurology 2002 Dec 24;59(12):1944-1950 [FREE Full text] [Medline: 12499488]
13. Hinkin CH, Hardy DJ, Mason KI, Castellon SA, Durvasula RS, Lam MN, et al. Medication adherence in HIV-infected adults: effect of patient age, cognitive status, and substance abuse. AIDS 2004 Jan 01;18 Suppl 1:S19-S25 [FREE Full text] [Medline: 15075494]
14. Vivithanaporn P, Heo G, Gamble J, Krentz HB, Hoke A, Gill MJ, et al. Neurologic disease burden in treated HIV/AIDS predicts survival: a population-based study. Neurology 2010 Sep 28;75(13):1150-1158 [FREE Full text] [doi: 10.1212/WNL.0b013e3181f4d5bb] [Medline: 20739646]
15. Ettenhofer ML, Foley J, Castellon SA, Hinkin CH. Reciprocal prediction of medication adherence and neurocognition in HIV/AIDS. Neurology 2010 Apr 13;74(15):1217-1222 [FREE Full text] [doi: 10.1212/WNL.0b013e3181d8c1ca] [Medline: 20220123]
16. Ettenhofer ML, Hinkin CH, Castellon SA, Durvasula R, Ullman J, Lam M, et al. Aging, neurocognition, and medication adherence in HIV infection. Am J Geriatr Psychiatry 2009 Apr;17(4):281-290 [FREE Full text] [doi: 10.1097/JGP.0b013e31819431bd] [Medline: 19307857]
17. Anand P, Springer SA, Copenhaver MM, Altice FL. Neurocognitive impairment and HIV risk factors: a reciprocal relationship. AIDS Behav 2010 Dec;14(6):1213-1226 [FREE Full text] [doi: 10.1007/s10461-010-9684-1] [Medline: 20232242]
18. Iudicello JE, Woods SP, Cattie JE, Doyle K, Grant I, HIV Neurobehavioral Research Program Group. Risky decision-making in HIV-associated neurocognitive disorders (HAND). Clin Neuropsychol 2013;27(2):256-275 [FREE Full text] [doi: 10.1080/13854046.2012.740077] [Medline: 23181946]
19. Thames AD, Streiff V, Patel SM, Panos SE, Castellon SA, Hinkin CH. The role of HIV infection, cognition, and depression in risky decision-making. J Neuropsychiatry Clin Neurosci 2012;24(3):340-348 [FREE Full text] [doi: 10.1176/appi.neuropsych.11110340] [Medline: 23037648]

20. Gorman AA, Foley JM, Ettenhofer ML, Hinkin CH, van GWG. Functional consequences of HIV-associated neuropsychological impairment. Neuropsychol Rev 2009 Jun;19(2):186-203 [FREE Full text] [doi: 10.1007/s11065-009-9095-0] [Medline: 19472057]
21. Kamminga J, Cysique LA, Lu G, Batchelor J, Brew BJ. Validity of cognitive screens for HIV-associated neurocognitive disorder: a systematic review and an informed screen selection guide. Curr HIV/AIDS Rep 2013 Dec;10(4):342-355 [FREE Full text] [doi: 10.1007/s11904-013-0176-6] [Medline: 24072534]
22. Mind Exchange Working Group. Assessment, diagnosis, and treatment of HIV-associated neurocognitive disorder: a consensus report of the mind exchange program. Clin Infect Dis 2013 Apr;56(7):1004-1017 [FREE Full text] [doi: 10.1093/cid/cis975] [Medline: 23175555]
23. Cysique L, Bain M, Lane T, Brew B. Management issues in HIV-associated neurocognitive disorders. NBHIV 2012 Jul;4:63-73. [doi: 10.2147/NBHIV.S30466]
24. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global Report: UNAIDS Report on the Global AIDS Epidemic 2013. Geneva, Switzerland: UNAIDS; 2013.
25. World Health Statistics: Global Health Indicators. Geneva, Switzerland: World Health Organization; 2010.
26. Heaton RK, Marcotte TD, Mindt MR, Sadek J, Moore DJ, Bentley H, HNRC Group. The impact of HIV-associated neuropsychological impairment on everyday functioning. J Int Neuropsychol Soc 2004 May;10(3):317-331. [doi: 10.1017/S1355617704102130] [Medline: 15147590]
27. Robbins RN, Remien RH, Mellins CA, Joska JA, Stein DJ. Screening for HIV-associated dementia in South Africa: potentials and pitfalls of task-shifting. AIDS Patient Care STDS 2011 Oct;25(10):587-593 [FREE Full text] [doi: 10.1089/apc.2011.0154] [Medline: 21919734]
28. Breuer E, Stoloff K, Myer L, Seedat S, Stein DJ, Joska J. Reliability of the lay adherence counsellor administered substance abuse and mental illness symptoms screener (SAMISS) and the International HIV Dementia Scale (IHDS) in a primary care HIV clinic in Cape Town, South Africa. AIDS Behav 2012 Aug;16(6):1464-1471. [doi: 10.1007/s10461-011-0067-z] [Medline: 22001934]
29. Zipursky AR, Gogolishvili D, Rueda S, Brunetta J, Carvalhal A, McCombe JA, et al. Evaluation of brief screening tools for neurocognitive impairment in HIV/AIDS: a systematic review of the literature. AIDS 2013 Sep 24;27(15):2385-2401 [FREE Full text] [doi: 10.1097/QAD.0b013e328363bf56] [Medline: 23751261]
30. Joska JA, Westgarth-Taylor J, Hoare J, Thomas KG, Paul R, Myer L, et al. Validity of the International HIV Dementia Scale in South Africa. AIDS Patient Care STDS 2011 Feb;25(2):95-101. [doi: 10.1089/apc.2010.0292] [Medline: 21214343]
31. Joska JA, Witten J, Thomas KG, Robertson C, Casson-Crook M, Roosa H, et al. A comparison of five brief screening tools for HIV-associated neurocognitive disorders in the USA and South Africa. AIDS Behav 2016 Aug;20(8):1621-1631. [doi: 10.1007/s10461-016-1316-y] [Medline: 26860536]
32. Callaghan M, Ford N, Schneider H. A systematic review of task- shifting for HIV treatment and care in Africa. Hum Resour Health 2010 Mar 31;8:8. [doi: 10.1186/1478-4491-8-8] [Medline: 20356363]
33. Bekker L, Myer L, Orrell C, Lawn S, Wood R. Rapid scale-up of a community-based HIV treatment service: programme performance over 3 consecutive years in Guguletu, South Africa. S Afr Med J 2006 Apr;96(4):315-320 [FREE Full text] [Medline: 16670804]
34. Zachariah R, Ford N, Philips M, Lynch S, Massaquoi M, Janssens V, et al. Task shifting in HIV/AIDS: opportunities, challenges and proposed actions for sub-Saharan Africa. Trans R Soc Trop Med Hyg 2009 Jun;103(6):549-558. [doi: 10.1016/j.trstmh.2008.09.019] [Medline: 18992905]
35. Robbins RN, Brown H, Ehlers A, Joska JA, Thomas KG, Burgess R, et al. A smartphone app to screen for HIV-related neurocognitive impairment. J Mob Technol Med 2014 Feb 01;3(1):23-26 [FREE Full text] [doi: 10.7309/jmtm.3.1.5] [Medline: 24860624]
36. Robbins RN, Mellins CA, Leu C, Rowe J, Warne P, Abrams EJ, et al. Enhancing lay counselor capacity to improve patient outcomes with multimedia technology. AIDS Behav 2015 Jun;19 Suppl 2:163-176. [doi: 10.1007/s10461-014-0988-4] [Medline: 25566763]
37. Remien RH, Mellins CA, Robbins RN, Kelsey R, Rowe J, Warne P, et al. Masivukeni: development of a multimedia based antiretroviral therapy adherence intervention for counselors and patients in South Africa. AIDS Behav 2013 Jul;17(6):1979-1991 [FREE Full text] [doi: 10.1007/s10461-013-0438-8] [Medline: 23468079]
38. Gouse H, Henry M, Robbins RN, Lopez-Rios J, Mellins CA, Remien RH, et al. Psychosocial aspects of ART counseling: a comparison of HIV beliefs and knowledge in PMTCT and ART-naive women. J Assoc Nurses AIDS Care 2017;28(4):504-517. [doi: 10.1016/j.jana.2017.03.002] [Medline: 28377125]
39. Benedict RH, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test-Revised: normative data and analysis of inter-form and test-retest reliability. Clinical Neuropsychol 1998 Feb 1;12(1):43-55. [doi: 10.1076/clin.12.1.43.1726]
40. Benedict RH. Brief Visuospatial Memory Test-Revised: Professional Manual. Odessa, FL: Psychological Assessment Resources, Inc; 1997.
41. D'Elia L, Satz P, Uchiyama C, White T. Color Trails Test. Odessa, FL: Psychological Assessment Resources, Inc; 1996.
42. Kongs S, Thompson L, Iverson G, Heaton R. Wisconsin Card Sorting Test-64 Card Version: Professional Manual. Odessa, FL: Psychological Assessment Resources; 2000.

43. Wechsler D. WMS-III: Wechsler Memory Scale Administration and Scoring Manual. San Antonio, TX: Psychological Corporation; 1997.
44. Wechsler D. WAIS-III: Wechsler Adult Intelligence Scale Administration and Scoring Manual. San Antonio, TX: Psychological Corporation; 1997.
45. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. PMS 1958 Dec;8(7):271. [doi: 10.2466/PMS.8.7.271-276]
46. Tombaugh TN, Kozak J, Rees L. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. Arch Clin Neuropsychol 1999 Feb;14(2):167-177. [Medline: 14590600]
47. Strauss E, Sherman E, Spreen O. A compendium of neuropsychological tests: administration, norms, and commentary. Oxford: Oxford University Press; 2006.
48. Heaton RK, Clifford DB, Franklin DR, Woods SP, Ake C, Vaida F, CHARTER Group. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. Neurology 2010 Dec 07;75(23):2087-2096 [FREE Full text] [doi: 10.1212/WNL.0b013e318200d727] [Medline: 21135382]
49. Carey CL, Woods SP, Gonzalez R, Conover E, Marcotte TD, Grant I, HNRC Group. Predictive validity of global deficit scores in detecting neuropsychological impairment in HIV infection. J Clin Exp Neuropsychol 2004 May;26(3):307-319. [doi: 10.1080/13803390490510031] [Medline: 15512922]
50. Blackstone K, Moore DJ, Heaton RK, Franklin DR, Woods SP, Clifford DB, CNS HIV Antiretroviral Therapy Effects Research (CHARTER) Group. Diagnosing symptomatic HIV-associated neurocognitive disorders: self-report versus performance-based assessment of everyday functioning. J Int Neuropsychol Soc 2012 Jan;18(1):79-88 [FREE Full text] [doi: 10.1017/S135561771100141X] [Medline: 22114912]
51. Carey C, Woods S, Rippeth J, Gonzalez R, Moore D, Marcotte T, HNRC Group. Initial validation of a screening battery for the detection of HIV-associated cognitive impairment. Clin Neuropsychol 2004 May;18(2):234-248. [doi: 10.1080/13854040490501448] [Medline: 15587671]

Abbreviations

ART: antiretroviral therapy
AUC: area under the curve
BVMT-R: Brief Visuospatial Memory Test-Revised
GDS: global deficit score
HAND: HIV-associated neurocognitive disorder
HVLT-R: Hopkins Verbal Learning Test-Revised
NCI: neurocognitive impairment
NPV: negative predictive value
PLHIV: people living with HIV
PPV: positive predictive value
RCT: randomized controlled trial
ROC: receiver operator characteristic
WAIS-III: Wechsler Adult Intelligence Scales, Third Edition
WCST: Wisconsin Card Sorting Test
WMS-III: Wechsler Memory Scales, Third Edition

Edited by G Eysenbach; submitted 11.10.17; peer-reviewed by J Buttolph; comments to author 03.11.17; revised version received 21.11.17; accepted 23.11.17; published 05.01.18.

Please cite as:
Robbins RN, Gouse H, Brown HG, Ehlers A, Scott TM, Leu CS, Remien RH, Mellins CA, Joska JA
A Mobile App to Screen for Neurocognitive Impairment: Preliminary Validation of NeuroScreen Among HIV-Infected South African Adults
JMIR Mhealth Uhealth 2018;6(1):e5
URL: http://mhealth.jmir.org/2018/1/e5/
doi:10.2196/mhealth.9148
PMID:29305338

Original Paper

Toward mHealth Brief Contact Interventions in Suicide Prevention: Case Series From the Suicide Intervention Assisted by Messages (SIAM) Randomized Controlled Trial

Sofian Berrouiguet[1,2], MD; Mark Erik Larsen[3*], DPhil; Catherine Mesmeur[4]; Michel Gravey[5]; Romain Billot[1,2], PhD; Michel Walter[4], MD, PhD; HUGOPSY Network[6*]; Christophe Lemey[4], MD; Philippe Lenca[1], PhD

[1]IMT Atlantique, Lab-STICC, Université Bretagne Loire F-29238 Brest, Brest, France
[2]EA 7479 SPURBO, Université de Bretagne occidentale, Brest, France
[3]Black Dog Institute, University of New South Wales, Sydney, Australia
[4]Adult Psychiatry, Brest Medical University Hospital at Bohars, Brest, France
[5]Sys.Vision, Lannion, France
[6]Réseau HUGOPSY, Rennes, France
[*]these authors contributed equally

Corresponding Author:
Sofian Berrouiguet, MD
IMT Atlantique
Lab-STICC
Université Bretagne Loire F-29238 Brest
Brest Medical University Hospital at Bohars, Adult Psychiatry
Route de Ploudalmezeau
Brest, 29820
France
Phone: 33 668204178
Email: sofian.berrouiguet@gmail.com

Abstract

Background: Research indicates that maintaining contact either via letter or postcard with at-risk adults following discharge from care services after a suicide attempt (SA) can reduce reattempt risk. Pilot studies have demonstrated that interventions using mobile health (mHealth) technologies are feasible in a suicide prevention setting.

Objective: The aim of this study was to report three cases of patients recruited in the Suicide Intervention Assisted by Messages (SIAM) study to describe how a mobile intervention may influence follow-up.

Methods: SIAM is a 2-year, multicenter randomized controlled trial conducted by the Brest University Hospital, France. Participants in the intervention group receive SIAM text messages 48 hours after discharge, then at day 8 and day 15, and months 1, 2, 3, 4, 5, and 6. The study includes participants aged 18 years or older, who have attended a participating hospital for an SA, and have been discharged from the emergency department (ED) or a psychiatric unit (PU) for a stay of less than 7 days. Eligible participants are randomized between the SIAM intervention messages and a control group. In this study, we present three cases from the ongoing SIAM study that demonstrate the capability of a mobile-based brief contact intervention for triggering patient-initiated contact with a crisis support team at various time points throughout the mobile-based follow-up period.

Results: Out of the 244 patients recruited in the SIAM randomized controlled trial, three cases were selected to illustrate the impact of mHealth on suicide risk management. Participants initiated contact with the emergency crisis support service after receiving text messages up to 6 months following discharge from the hospital. Contact was initiated immediately following receipt of a text message or up to 6 days following a message.

Conclusions: This text message–based brief contact intervention has demonstrated the potential to reconnect suicidal individuals with crisis support services while they are experiencing suicidal ideation as well as in a period after receiving messages. As follow-up phone calls over an extended period of time may not be feasible, this intervention has the potential to offer simple technological support for individuals following discharge from the ED.

Trial Registration: ClinicalTrials.gov NCT02106949; https://clinicaltrials.gov/ct2/show/NCT02106949 (Archived by WebCite at http://www.webcitation.org/6wMtAFL49)

(JMIR Mhealth Uhealth 2018;6(1):e8) doi:10.2196/mhealth.7780

KEYWORDS

Suicide; Text Messaging; Electronic Health Records; Cell Phone; Secondary Prevention; Tertiary Prevention

Introduction

Brief Contact Intervention and Suicide Prevention

A previous suicide attempt (SA) is a strong predictor of death from future suicidal behaviors. Approximately one-third of individuals who attempt suicide seek treatment for their injuries from hospital emergency departments (EDs) [1], and the immediate period following discharge from hospital is critical for emergency and mental health care service follow-up, as most suicide reattempts occur within the first month of discharge [2]. For example, Hunt et al showed that 47% of fatal suicide reattempts occurred before the first scheduled follow-up appointment [3].

There has, therefore, been growing interest in the development of brief contact interventions (BCIs) delivered following discharge from the ED after an SA. BCIs are low-resource, nonintrusive interventions seeking to maintain long-term contact with patients after an SA. BCIs follow a structured schedule and remain operational over a sustained period of time. They commonly use short letters [4], postcards [5], phone calls [6], and crisis cards [7] to keep in contact with participants, without the provision of additional therapies. BCIs have been mostly used with clinical populations following presentation to an ED for self-harm, self-injury, self-poisoning or an SA. The content of BCIs differs between studies, but generally involves a short sentence expressing concern for the patient and emphasizing the availability of help should it be needed. BCIs have shown mixed or inconclusive results, but they indicate trends toward preventive effects in specific at-risk subgroups (eg, first suicide attempters, females, young suicide attempters) depending on the BCI employed. The pioneer intervention was proposed by Motto et al [4] and was based on postal contact. Motto provided 5 years of postal contact with patients who refused follow up after an SA. The objective was to show the patients that someone was concerned about their situation and maintain positive feelings toward them. After 5 years, a significant decrease in suicide-related deaths was observed in the contacted group (vs no-contact group). In a study by Carter et al [5], postcards were sent in the year following an SA, with a lower number of repeat episodes in the contacted group, especially among women. In general, these interventions aim to improve help-seeking and may facilitate access to health care services in the case of recurrence of suicidal ideation.

Mobile Health and Suicide Prevention

By the end 2017, mobile cellular subscriptions worldwide are expected to reach 4.3 billion globally, and mobile broadband subscriptions have grown more than 20% annually in the last 5 years. Mobile phones are generally kept on at all times and carried everywhere, making them an ideal platform for the broad implementation of personalized and mobile health (mHealth) interventions [8]. mHealth has the potential to reduce waiting times for appointments and reduce the need to meet in-person with a clinician, successively diminishing the workload of mental health professionals; to be more cost-effective to practices; and to encourage self-care strategies [9]. A substantial amount of interactive and psychoeducational apps are readily available to download concerning a wide range of health issues, including suicide prevention [10]. A recent review of existing technology-enhanced interventions addressed determinants of suicidal behavior [9]. Included studies examined the use of standalone or, in most cases, adjunctive technology-enhanced interventions for suicide prevention delivered by mobile phone app, text message, telephone, computer, Web, CD-ROM, and video.

Mobile phones have many characteristics that make them well suited for health interventions. For example, the frequent use of mobile phones is associated with an opportunity for mass communication. mHealth can be defined as *the use of mobile computing and communication technologies in health care and public health* [11]. mHealth interventions have the potential to incorporate qualities often associated with more effective health communication interventions, such as personalization, tailoring, interactivity, and message repetition at a relatively low cost. In the suicide prevention setting, mobile phone technology can support the transition of care, specific treatment targets, and safety planning—all of which are important elements of treatment [12,13].

Text Messages as a Suicide Prevention Brief Contact Intervention

Mobile text messaging (short message service, SMS) in particular has proven to be an effective form of psychiatric intervention [14]. Text messages can be sent in a standardized or individualized format and are available on all cellular phones, including low-cost devices. They can also be sent from a server-based platform that allows automatic prescheduling of message delivery and monitoring of delivery receipts. Studies such as Suicide Intervention Assisted by Messages (SIAM) [15] and Reconnecting AFTer a suicide attempt (RAFT) (Larsen et al, unpublished data, 2017) incorporate the use of stand-alone (RAFT) or adjunctive (SIAM) mobile phone-enhanced interventions for suicide prevention. The aim of these studies is to help connect participants with support services following discharge using SMS contacts, reducing repeat episodes of self-harm, reducing representations to the ED, and ultimately reducing deaths by suicide.

We conducted a feasibility and acceptability study of SIAM [16], demonstrating that the intervention was technically robust and well accepted by patients. A randomized controlled trial is

ongoing; however, we hypothesize that a descriptive analysis of selected cases would bring insight regarding the capability of a mobile-based BCI to trigger patient-initiated contact with emergency services. This case series aimed to identify cases of patients recruited in the SIAM study that may demonstrate the capability of a mobile-based brief contact intervention for triggering patient-initiated contact with a crisis support team at various time points throughout the mobile-based follow-up period.

Methods

Study Design

We performed a descriptive analysis on a selected sample of patients randomized in the intervention group of the SIAM study, recruited between August 24, 2014, and June 5, 2017. SIAM is a 2-year, multicentre randomized controlled trial conducted by the Brest University Hospital, France. The study was registered on Clinical Trials Registry (clinicaltrials.gov NCT02106949). The study protocol is described in detail elsewhere (see Multimedia Appendix 1) [15]. The study includes participants aged 18 years or over, who have attended a participating hospital for an SA, and have been discharged from the ED or a psychiatric unit (PU) for a stay of less than 7 days. Eligible participants are randomized between the SIAM intervention messages and a control group. This study was approved by the French West VI ethics committee.

Patients

Participants in the intervention group receive the SIAM text messages 48 hours after discharge, then at day 8 and day 15, and months 1, 2, 3, 4, 5, and 6. The messages refer to validation of the suffering, recall of the discharge agreement, and ongoing outreach care. The messages are also personalized with the participant's name, the monitoring doctor's name, and a local crisis telephone number. An example message contains the following text: "Mr X, we hope that your situation is getting better and that you can attend the consultation with Dr Y (April 7th 2011 at 10:00 h). You can call us for anything you may need at 0298000000."

All participants, regardless of their randomization, received treatment as usual (TAU) including a postdischarge consultation with a psychiatrist. Study monitoring phone calls by a trained psychiatrist were also conducted at 6 and 13 months to perform follow-up evaluations.

The primary outcome measure of the SIAM study is the number of participants who have a repeat SA within 6 months. Secondary outcomes assess the number of attempts after 13 months and the number of deaths at both time points. To date, 387 participants have been recruited to the SIAM study. As recruitment is ongoing, complete data on the number of crisis calls initiated by participants in the intervention arm are not available. Nevertheless, representative cases were identified by the trial management office to reflect contacts initiated by participants at different points through the study.

Inclusion criteria for this case series were as follows: patients recruited in the SIAM study who had attended both the 6-month and 13-month monitoring interviews, and who had contact with emergency services during the study period.

Members of the study office (SB, CM, LC, and MW) established an initial list of patients meeting the inclusion criteria. The contents of the monitoring interviews were reviewed to identify patients who had contact with emergency services during the study period. SB and MC finally selected patients who demonstrated the capability of a mobile-based brief contact intervention for triggering patient-initiated contact with a crisis support team at various points throughout the follow-up period.

Three cases were identified, and we report the assessment of the participants' baseline Mini International Neuropsychiatric Interview (MINI) [17] and Columbia Suicide Severity Rating Scale (CSSR-S) [18] and a narrative description of circumstances associated with their participant-initiated contact with the crisis team.

Results

Cases from the SIAM randomized controlled trial were identified to illustrate the impact of the mHealth brief contact intervention on further suicide ideation. The patient selection process for the case series is presented in Figure 1. Sociological main features of selected cases and history of SA are shown in Table 1. Psychiatric diagnoses are shown in Table 2.

Case 1 was recruited to the study after an SA by deliberate poisoning. He had a history of previous SAs 2 and 6 years previously, also by deliberate self-poisoning (Table 1). He was diagnosed with general anxiety disorder and dysthymia (Table 2). He had no family history of mental disorder. He was divorced and unemployed. He was randomized in the intervention group. He also had acute alcohol intoxication (alcohol 2g/L) and a psychiatric history of hospitalization for alcohol withdrawal. This SA occurred 7 months after receiving a diagnosis of a chronic health condition (major life event) and 1 month after losing his job. He clearly expressed suicide ideation with a wish to die before the deliberate poisoning but not when we performed baseline assessment (Table 3). After discharge, an appointment with his general physician (GP) was scheduled.

The patient called the emergency service seeking help for suicide ideation. He told the nurse he had the emergency phone number from the last SMS he received (the second SMS received 3 days previously). After the phone call, he attended the appointment at the ED suggested by the nurse.

Figure 1. Flow of patient selection process.

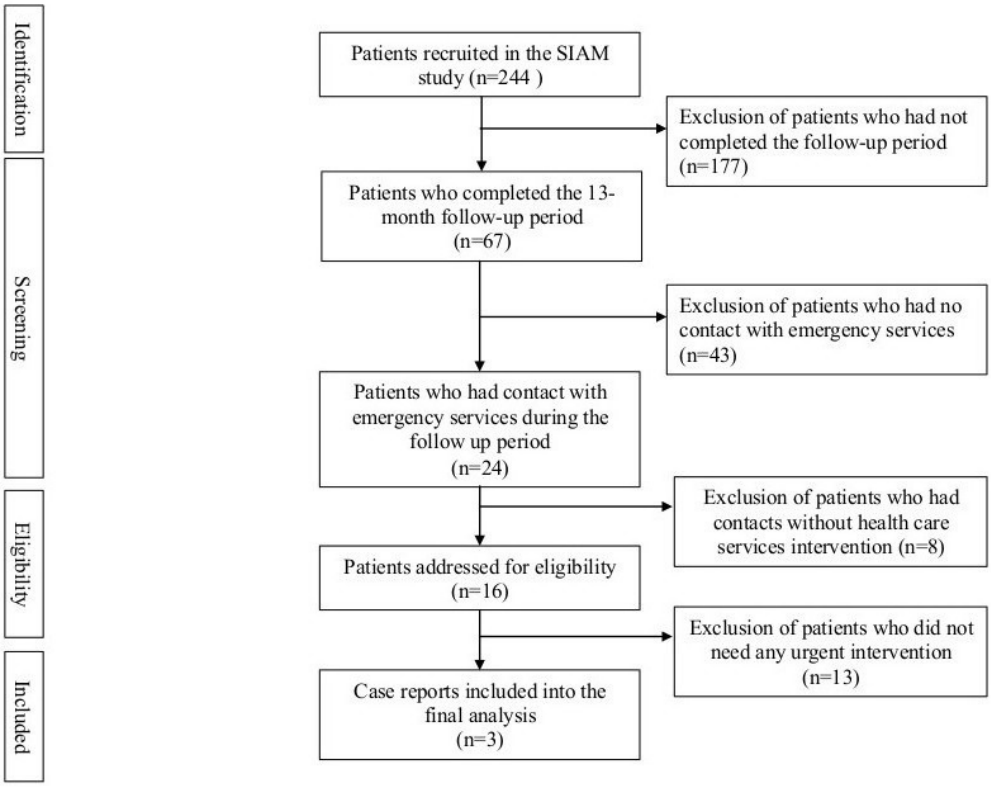


Table 1. Sociological main features and history of suicide attempt.

Patient	Sex	Age, years	Marital status	Employment	Total number of suicide attempts in lifetime	Number of suicide attempts in the last 3 years	Family history of mental disorder	Major life event
Case 1	Male	40-45	Divorced	Unemployed	2	1	No	Yes
Case 2	Male	55-60	In a relationship	Unemployed	4	2	Yes	No
Case 3	Female	40-45	In a relationship	Unemployed	5	3	No	No

Table 2. Psychiatric diagnosis of selected patients.

Patient	DSM[a] diagnosis	Alcohol dependence and abuse
Case 1	Generalized anxiety disorder	Yes (both)
Case 2	Dysthymia	Yes (both)
Case 3	Dysthymia, generalized anxiety disorder	No (both)

[a]DSM: Diagnostic and Statistical Manual.

Table 3. Characteristics of the most recent suicide attempt.

Patient	Wish to be dead	How many times have you had suicidal thoughts?	When you have suicidal thoughts how long do they last?	Has there been a time when you started to do something to end your life but someone or something stopped you before you actually did anything?
Case 1	No	—[a]	—	—
Case 2	Yes	Less than once a week	Fleeting few seconds or minutes	Yes
Case 3	Yes	Less than once a week	Less than 1 hour	Yes

[a]— indicates missing data.

Case 2 was recruited to the SIAM study after an SA by exsanguination. He was diagnosed with general anxiety disorder and dysthymia (Table 2). He was in a relationship and unemployed. He was randomized to the intervention group. The baseline evaluation indicated nonadherence to his medical treatment. He had a familial history of mental disorder (Table 1). The MINI assessment showed a diagnostic of dysthymia, alcohol abuse, and alcohol dependence (Table 2). The Columbia scale at baseline showed it was the fourth time he attempted suicide (Table 1). His most serious attempt occurred over 30 years ago when he was rescued from hanging. During the past few days, he had suicidal thoughts less than once a week. These thoughts usually lasted a few seconds (Table 3).

He received the ninth text message of the study and called the emergency service 6 days later. During the phone call, he disclosed suicidal ideation and that he had been drinking alcohol. He accepted the proposal to be driven to the ED by an emergency transport. He was subsequently admitted to a hospital for alcohol withdrawal.

Case 3 was recruited after an SA by deliberate self-poisoning. She had a history of 5 previous SAs (Table 1). She was in a relationship and unemployed. The most recent attempt occurred in the context of a conflicting relationship with her adolescent son. Before the SA, she had suicidal thoughts less than once a week. These thoughts usually lasted less than an hour (Table 3).

The MINI scale at baseline showed a personal history of dysthymia. The Columbia scale at baseline showed her first SA occurred 4 years ago. After discharge, an appointment with her GP was scheduled. The patient received the fifth text message on a day when she was experiencing intense suicidal ideation. Immediately after receiving the message, she called the phone number provided in the message. She got in touch with the emergency service that proposed an immediate intervention of an emergency transport. Within 1 hour of receiving the message, she was admitted to the emergency service.

Discussion

Principal Findings

We have presented three cases describing situations within the SIAM intervention where crisis support services offered through this BCI have been initiated. In each case, the contact has been initiated by the study participant immediately after receiving a message (Case 3) or a few days later (Case 1 and 2). These contacts have also been initiated over a range of periods since discharge, from 1 week (Case 1) to 6 months (Case 2). These cases highlight the potential for connecting individuals to crisis services after an SA using automated text messages. The use of text messaging is likely to be more cost-effective than attempting follow-up phone calls, which require considerable on-going resources that may limit the feasibility of regular follow-ups. Furthermore, our experience from other studies indicates that calls are frequently not answered [19].

The sociodemographic details in Table 1 indicate that all the 3 cases relate to individuals aged over 40 years, and all the 3 were unemployed. This is perhaps surprising, as mHealth interventions are often considered to have the greatest potential for younger participants. These results demonstrate the potential for such interventions across a range of sociodemographic characteristics.

This case series presents an example of a real-world intervention triggered by an mHealth intervention. In particular, this intervention aims to strengthen the connection between patients and their care team by encouraging contact during a crisis. Similar low-intensity interventions also have the potential to be made available to a wider population of participants, for example, through other health care services or self-registrations from individuals who may be in crisis but not otherwise seeking help. There is, therefore, a potential to connect more people with appropriate care before a hospital presentation.

Limitations

Although these cases demonstrate the feasibility of initiating crisis contact using text messages, the effectiveness cannot yet be determined as the SIAM study is ongoing. It is therefore too early to address whether a text message–based BCI can reduce repeat episodes and mortality. Furthermore, the comparator arm in this study is TAU; therefore, the effectiveness of a text message–based BCI versus a telephone-based BCI cannot be compared. Further investigation into the methods [20] and mechanisms [21] of BCIs is warranted.

The SIAM mobile intervention was proposed as an adjunct to existing treatment strategies and not as a substitute. It was not designed as a remote counseling service [22].

As a result, the text message–based contact was deliberately limited to a one-way communication. We disabled the feature of two-way communication to encourage phone calls and face-to-face contacts. However, as illustrated, a simple message can have an important impact. Contents of messages are of particular importance, and certain key characteristics such as personalization, caring sentiment, and polite text are associated with more successful preventative messages [23].

As with any outreach activity, this intervention presents a possible risk of intrusiveness into the daily lives of participants. This issue has not been assessed in the papers we reviewed or in other reviews in the field [24], and in our earlier feasibility study, no patients reported finding this mHealth intervention intrusive [16]. Conversely, in the third case reported here, a text message was received at a timely moment. However, an SMS may also arrive at an inconvenient time; therefore, informational messages may be ignored or deleted. Extended interventions incorporating momentary assessments may additionally lack responses at these times. Researchers should ensure that such burden would not be detrimental to participants' well-being, particularly when studying individuals who have recently been discharged or may currently be facing a crisis.

In our study, the SMS intervention was proposed as an adjunct of TAU. Others have suggested additional supportive outreach and coping strategies; for example, in the Brief Mobile Treatment (BMT) intervention [25], a patient received generic weekly text messages up to 26 weeks, supportive phone calls, and access to audio phone messages to reinforce psychotherapy principles. BMT participants received text reminders about

meditation, problem solving, spiritual or philosophical ideas, the importance of social support, avoiding alcohol and drugs, and crisis helpline details. Our eventual aim is to integrate an mHealth intervention within existing emergency care procedures, as this may increase the effectiveness of both the mHealth intervention and emergency services.

The Web app we developed for the SIAM study allowed the patient to respond to messages they received. However, as part of our safety protocol to avoid the possibility of unanswered crisis messages, we disabled the two-way communication feature. Since we started the study, natural language processing (NLP) machine learning prediction methods have been shown to predict suicide risk as well as heightened psychiatric symptoms in free-text responses sent via a mobile phone [26]. NLP methods applied on SMS communication may help to create low-cost and effective alternatives to traditional resource-heavy data monitoring systems and support decision making of clinicians.

SMS is a powerful tool to connect the patient with health care services. Being a simple and affordable technology, it also allows for the addition of other content as links to support websites and self-monitoring apps. A recent review showed that mHealth interventions had a larger effect when used for conveying psychoeducation. Safety planning, for example, can be easily implemented [27]. Furthermore, assessment of symptoms could also be performed using SMS, and it may allow for an accurate assessment of suicide ideations [28]. These features may help to identify, together with face-to-face assessment, personalized trajectories of symptoms, cognitive abilities, and symptoms impacting other symptoms [29].

The use of SMS allows the health care team to keep in touch with people that are not reachable by existing BCIs. It may be used as a useful additional or alternative to standard care when people decline other forms of treatment. Although we used a Web system to deploy our intervention, our main goal is to encourage person-to-person contact between suicide attempters and health care providers. Torous et al [30] evaluated an mHealth intervention that successfully encouraged participants to be more open when disclosing symptoms of depression. We believe that our mHealth intervention may also improve the likelihood of patients disclosing suicide ideation to emergency services and seeking help. The full effects of the SIAM intervention and related results on suicide reattempt reduction will be reported upon conclusion of the trial.

Conclusions

Brief contact interventions are a promising technique for maintaining contact with patients following discharge from the ED and preventing repeat SAs. This case series has demonstrated the ability of text message BCIs to encourage patients to contact health care services in times of a crisis and over periods up to 6 months following discharge from the hospital. An ongoing randomized controlled trial of the SIAM intervention aims to demonstrate the effectiveness of such mHealth BCIs for suicide prevention.

Acknowledgments

The authors would like to acknowledge the Lanestel company and are grateful to Sara Brandt for her contribution in writing this paper. This paper is funded by a grant from the Fondation de France, CRITT santé Bretagne, Fondation de l'avenir and Harmony Mutuelle, the French Embassy in Madrid. MEL was supported by a Society of Mental Health 2015 Early Career Research Award and a FASIC Fellowship from the Australian Academy of Science and the French Embassy in Australia.

Conflicts of Interest

None declared.

Authors' Contributions

Authors SB, MW, and CM were involved in the conception of the study and made substantial contributions to the study design. MG programmed the SMS monitoring software system. In addition, RB and PL made significant contributions to acquisition of data, training, and clinical supervision. All authors contributed to the analysis and interpretation of the data, were involved in drafting the manuscript, and have approved the final version.

Multimedia Appendix 1

Siam protocol description video.

[M4V File, 2MB - mhealth_v6i1e8_app1.m4v]

Multimedia Appendix 2

CONSORT checklist.

[PDF File (Adobe PDF File), 677KB - mhealth_v6i1e8_app2.pdf]

References

1. Finkelstein Y, Macdonald EM, Hollands S, Sivilotti ML, Hutson JR, Mamdani MM, Canadian Drug Safety and Effectiveness Research Network (CDSERN). Risk of Suicide Following Deliberate Self-poisoning. JAMA Psychiatry 2015 Jun;72(6):570-575. [doi: 10.1001/jamapsychiatry.2014.3188] [Medline: 25830811]
2. Meehan J, Kapur N, Hunt IM, Turnbull P, Robinson J, Bickley H, et al. Suicide in mental health in-patients and within 3 months of discharge. National clinical survey. Br J Psychiatry 2006 Feb;188:129-134 [FREE Full text] [doi: 10.1192/bjp.188.2.129] [Medline: 16449699]
3. Hunt IM, Kapur N, Webb R, Robinson J, Burns J, Shaw J, et al. Suicide in recently discharged psychiatric patients: a case-control study. Psychol Med 2009 Mar;39(3):443-449. [doi: 10.1017/S0033291708003644] [Medline: 18507877]
4. Motto JA, Bostrom AG. A randomized controlled trial of postcrisis suicide prevention. Psychiatr Serv 2001 Jun;52(6):828-833. [Medline: 11376235]
5. Carter GL, Clover K, Whyte IM, Dawson AH, D'Este C. Postcards from the EDge project: randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. Br Med J 2005 Oct 08;331(7520):805 [FREE Full text] [doi: 10.1136/bmj.38579.455266.E0] [Medline: 16183654]
6. Vaiva G, Vaiva G, Ducrocq F, Meyer P, Mathieu D, Philippe A, et al. Effect of telephone contact on further suicide attempts in patients discharged from an emergency department: randomised controlled study. Br Med J 2006 May 27;332(7552):1241-1245 [FREE Full text] [doi: 10.1136/bmj.332.7552.1241] [Medline: 16735333]
7. Evans J, Evans M, Morgan HG, Hayward A, Gunnell D. Crisis card following self-harm: 12-month follow-up of a randomised controlled trial. Br J Psychiatry 2005 Aug;187:186-187 [FREE Full text] [doi: 10.1192/bjp.187.2.186] [Medline: 16055834]
8. itu.int. Facts and figures 2017 URL: http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [accessed 2017-08-29] [WebCite Cache ID 6vBezidyV]
9. Kreuze E, Jenkins C, Gregoski M, York J, Mueller M, Lamis DA, et al. Technology-enhanced suicide prevention interventions: a systematic review. J Telemed Telecare 2017 Jul;23(6):605-617. [doi: 10.1177/1357633X16657928] [Medline: 27377792]
10. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
11. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
12. Skovgaard LJL, Frandsen H, Erlangsen A. MYPLAN - A mobile phone application for supporting people at risk of suicide. Crisis 2016 May;37(3):236-240. [doi: 10.1027/0227-5910/a000371] [Medline: 26831213]
13. Kennard BD, Biernesser C, Wolfe KL, Foxwell AA, Craddock LS, Rial KV, et al. Developing a brief suicide prevention intervention and mobile phone application: a qualitative report. J Technol Hum Serv 2015 Oct 01;33(4):345-357 [FREE Full text] [doi: 10.1080/15228835.2015.1106384] [Medline: 26977137]
14. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
15. Berrouiguet S, Alavi Z, Vaiva G, Courtet P, Baca-García E, Vidailhet P, et al. SIAM (Suicide intervention assisted by messages): the development of a post-acute crisis text messaging outreach for suicide prevention. BMC Psychiatry 2014;14:294 [FREE Full text] [doi: 10.1186/s12888-014-0294-8] [Medline: 25404215]
16. Berrouiguet S, Gravey M, Le GM, Alavi Z, Walter M. Post-acute crisis text messaging outreach for suicide prevention: a pilot study. Psychiatry Res 2014 Jul 30;217(3):154-157. [doi: 10.1016/j.psychres.2014.02.034] [Medline: 24736112]
17. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34. [Medline: 9881538]
18. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
19. Larsen ME, Curry L, Mastellos N, Robb C, Car J, Middleton LT. Development of the CHARIOT research register for the prevention of Alzheimer's dementia and other late onset neurodegenerative diseases. PLoS One 2015;10(11):e0141806 [FREE Full text] [doi: 10.1371/journal.pone.0141806] [Medline: 26599631]
20. Vaiva G, Walter M, Al Arab AS, Courtet P, Bellivier F, Demarty AL, et al. ALGOS: the development of a randomized controlled trial testing a case management algorithm designed to reduce suicide risk among suicide attempters. BMC Psychiatry 2011 Jan 02;11:1 [FREE Full text] [doi: 10.1186/1471-244X-11-1] [Medline: 21194496]
21. Milner A, Carter G, Pirkis J, Robinson J, Spittal MJ. Letters, green cards, telephone calls and postcardsystematic and meta-analytic review of brief contact interventions for reducing self-harm, suicide attempts and suicide. Br J Psychiatry 2015 Mar 15;206(3):184-190. [doi: 10.1192/bjp.bp.114.147819] [Medline: 25733570]
22. Thiha P, Pisani AR, Gurditta K, Cherry E, Peterson DR, Kautz H, et al. Efficacy of web-based collection of strength-based testimonials for text message extension of youth suicide prevention program: randomized controlled experiment. JMIR Public Health Surveill 2016 Nov 09;2(2):e164 [FREE Full text] [doi: 10.2196/publichealth.6207] [Medline: 27829575]

23. Odeny TA, Newman M, Bukusi EA, McClelland RS, Cohen CR, Camlin CS. Developing content for a mHealth intervention to promote postpartum retention in prevention of mother-to-child HIV transmission programs and early infant diagnosis of HIV: a qualitative study. PLoS One 2014;9(9):e106383 [FREE Full text] [doi: 10.1371/journal.pone.0106383] [Medline: 25181408]
24. Mbuagbaw L, van der Kop ML, Lester RT, Thirumurthy H, Pop-Eleches C, Smieja M, et al. Mobile phone text messages for improving adherence to antiretroviral therapy (ART): a protocol for an individual patient data meta-analysis of randomised trials. BMJ Open 2013;3(5):e002954 [FREE Full text] [doi: 10.1136/bmjopen-2013-002954] [Medline: 23793704]
25. Marasinghe RB, Edirippulige S, Kavanagh D, Smith A, Jiffry MT. Effect of mobile phone-based psychotherapy in suicide prevention: a randomized controlled trial in Sri Lanka. J Telemed Telecare 2012 Apr;18(3):151-155. [doi: 10.1258/jtt.2012.SFT107] [Medline: 22362830]
26. Cook BL, Progovac AM, Chen P, Mullin B, Hou S, Baca-Garcia E. Novel use of Natural Language Processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid. Comput Math Methods Med 2016;2016:8708434 [FREE Full text] [doi: 10.1155/2016/8708434] [Medline: 27752278]
27. Boudreaux ED, Brown GK, Stanley B, Sadasivam RS, Camargo CA, Miller IW. Computer administered safety planning for individuals at risk for suicide: development and usability testing. J Med Internet Res 2017 May 15;19(5):e149 [FREE Full text] [doi: 10.2196/jmir.6816] [Medline: 28506957]
28. Husky M, Swendsen J, Ionita A, Jaussent I, Genty C, Courtet P. Predictors of daily life suicidal ideation in adults recently discharged after a serious suicide attempt: a pilot study. Psychiatry Res 2017 Oct;256:79-84. [doi: 10.1016/j.psychres.2017.06.035] [Medline: 28624676]
29. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
30. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the Patient Health Questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]

Abbreviations

BCIs: brief contact interventions
BMT: Brief Mobile Treatment
CSSR-S: Columbia Suicide Severity Rating Scale
ED: emergency department
GP: general physician
mHealth: mobile health
MINI: Mini International Neuropsychiatric Interview
NLP: natural language processing
PU: psychiatric unit
RAFT: Reconnecting AFTer a suicide attempt
SA: suicide attempt
SIAM: Suicide Intervention Assisted by Messages
SMS: short message service
TAU: treatment as usual

Edited by G Eysenbach; submitted 05.04.17; peer-reviewed by K Gurditta, N Bragazzi, TR Soron, R Shrestha; comments to author 12.07.17; revised version received 06.09.17; accepted 03.10.17; published 10.01.18.

Please cite as:
Berrouiguet S, Larsen ME, Mesmeur C, Gravey M, Billot R, Walter M, HUGOPSY Network, Lemey C, Lenca P
Toward mHealth Brief Contact Interventions in Suicide Prevention: Case Series From the Suicide Intervention Assisted by Messages (SIAM) Randomized Controlled Trial
JMIR Mhealth Uhealth 2018;6(1):e8
URL: http://mhealth.jmir.org/2018/1/e8/
doi:10.2196/mhealth.7780
PMID:29321126

Original Paper

Usage of an Exercise App in the Care for People With Osteoarthritis: User-Driven Exploratory Study

Dorthe Boe Danbjørg[1,2*], RN, MSc, PhD; Allan Villadsen[3*], MD, PhD; Ester Gill[4*], BMed; Mette Juel Rothmann[1,5,6*], RN, MHSc, PhD; Jane Clemensen[1,7*], RN, PhD

[1]Centre for Innovative Medical Technology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[2]Quality of Life Research Center, Department of Haematology, Odense University Hospital, Odense, Denmark

[3]Orthopaedic Research Unit, Department of Orthopaedic Surgery and Traumatology, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[4]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[5]Research Unit of Medical Endocrinology, Department of Endocrinology, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[6]Research Unit of Rheumatology, Department of Rheumatology, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[7]Research Unit of Paediatrics, Hans Christian Andersen Children's Hospital, University of Southern Denmark, Odense University Hospital, Odense, Denmark

[*]all authors contributed equally

Corresponding Author:
Dorthe Boe Danbjørg, RN, MSc, PhD
Centre for Innovative Medical Technology
Department of Clinical Research
University of Southern Denmark, Odense University Hospital
Sdr Boulevard 29
Odense, 5000
Denmark
Phone: 45 26202186
Fax: 45 26202186
Email: dortheboe@gmail.com

Abstract

Background: Exercise has proven to reduce pain and increase quality of life among people living with osteoarthritis (OA). However, one major challenge is adherence to exercise once supervision ends.

Objective: This study aimed to identify mental and physical barriers and motivational and social aspects of training at home, and to test or further develop an exercise app.

Methods: The study was inspired from participatory design, engaging users in the research process. Data were collected through focus groups and workshops, and analyzed by systematic text condensation.

Results: Three main themes were found: competition as motivation, training together, and barriers. The results revealed that the participants wanted to do their training and had knowledge on exercise and pain but found it hard to motivate themselves. They missed the observation, comments, and encouragement by the supervising physiotherapist as well as their peers. Ways to optimize the training app were identified during the workshops as participants shared their experience.

Conclusions: This study concludes that the long-term continuation of exercising for patients with OA could be improved with the use of a technology tailored to users' needs, including motivational and other behavioral factors.

(JMIR Mhealth Uhealth 2018;6(1):e11) doi:10.2196/mhealth.7734

KEYWORDS

arthritis; rehabilitation; telemedicine

Introduction

Background

Osteoarthritis (OA) is a degenerative joint disease that causes pain and decreases physical function and quality of life. It is the most common musculoskeletal disorder [1,2], and globally, it is a heavy economic burden [3] with annual costs of US $89.1 billion in the US alone [4].

Exercise reduces pain at the same level as simple analgesics and nonsteroidal anti-inflammatory drugs for people with OA of the hip and knee [5]. Furthermore, it increases physical function and quality of life [6]. Hence, exercise is considered one of the cornerstones in the treatment of hip and knee OA [6-9]. General recommendations are to offer OA patients information or education of the different aspects of the disease in conjunction with supervised exercise for 6 to 12 weeks [10,11]. Exercise is encouraged to be continued lifelong. In Scandinavia, this approach is generally accepted by both patients and health professionals but tends to be an underutilized treatment option among medical practitioners [12-14].

A dose-response relationship has been demonstrated between adherence to exercise and effect on people with knee OA [15]. Thus, the effectiveness of exercise on pain relief and disability only lasts as long as the patient participates [16,17]. However, when prescribed or allocated to exercise therapy, one major challenge seems to be adherence to exercise once supervision ends [18]. There may be many individual barriers to exercise—for example, busy daily schedules and lack of motivation. Accessible technologies may, to some extent, address these barriers. The percentage of the population who owns a smartphone or a desktop computer and has access to the Internet is rising and includes the elderly age group. Hence, an information technology solution may be able to reach a wide population. Furthermore, it may have economic benefits because treating more patients will not necessarily require more hours from physicians or other health care professionals [19].

Internet-Based Training

Internet-based training concepts aiming to improve the exercise level among chronically ill patients, such as patients with hip and knee OA, have already been developed and tested, but it is still essential to identify features in these exercise apps that will lead to sustained long-term usage [20,21]. An individualized approach to exercise is deemed essential for an optimal effect in OA treatment [18]. With the use of on a Web-based training app, it would be possible to personalize the exercise program and individualize motivational factors, which may optimize an individual's outcomes. Some studies show that class-based exercise is more effective in regard to adherence. Group interaction could potentially be added to the exercise app's interface, which generally is customized, to fit the personal needs of the individual user [5,21].

In 2014, an exercise app (Therapeutical exercise [Ther-ex]) (Ther-Ex APS, Denmark) targeting people with OA was brought to market in both iOS and Android versions. The developers were an orthopedic surgeon and a physiotherapist. Their concept for the app was to compile general OA recommendations of exercise and its monitoring into a solution, which was readily available for people with OA. The app contains approximately 100 individual videos of land-based functional exercises, which can be combined into individualized exercise programs (Textbox 1). Furthermore, the app contains an exercise and pain rating log, and these data can be displayed in various ways.

Aim

Limited user feedback on the app has generally been positive. However, troublesome functions have been identified when the app was tested in 2014. The current content was found insufficient to support adherence to exercise. To solve these problems and improve the app and the resultant self-care for people with OA, a systematic approach based on user participation was chosen.

The aim of this study was threefold:

1. To identify the mental and physical barriers and motivational and social aspects of training at home for people with the hip or knee arthritis
2. To test an exercise app for use at home by patients with hip or knee OA
3. To enhance the app on the basis of users' experiences with physical barriers, their motivational and social aspects of training, and their experiences with using the app

Textbox 1. Exercise app—therapeutical exercise.

Idea: to compile general osteoarthritis (OA) recommendations of exercise and its monitoring into an easy accessible tool

Contains: approximately 100 videos of land-based functional exercises and combines these into the following:

- Exercise programs
- Visual pain rating scale
- Exercise and weight diary
- Module for visualization of the above

Developed by a physiotherapist and an orthopedic surgeon with both clinical and research knowledge of OA

Methods

Design

This study was inspired from participatory design (PD) where the idea is to engage the users to innovate and develop technologies together with developers [22,23]. In a traditional PD project, the users are engaged from the beginning in defining the problem, which helps ensure that the technology meets the needs of the users. Thereafter, they are engaged in designing the technology and finally testing it [24]. In our case, we wanted to redesign the technology to meet the needs of the users, and therefore, we first explored how the users experienced problems related to their OA and rehabilitation. Then, the participants were invited to test and redesign by transforming the participants from merely informants to participants. They were asked to not just answer questions in an interview about their point of view but were also asked to actively participate in the testing and redesigning of the app, where the participants together with the researchers were *making* a mock-up of a future app [22]. The *making* of things can be a means of design participation, where the chosen tools used in the workshop allow the ability to create. The participants used their hands for expressing thoughts and ideas in the form of artifacts, which described the future app.

This study was inspired from hermeneutics philosophy, where the perspective has been to understand the participants' lived experiences in relation to living with OA to develop a technology that meets their needs. The interpretative approach focuses on understanding experiences and on how humans make sense of their subjective reality and attach meaning to it [25,26].

Sample and Context

Recruitment of Participants

The participants were recruited from Slagelse municipality (Denmark) where exercise is offered to people with OA via the Good Life with osteoArthritis in Denmark (GLA:D) project.

GLA:D is an initiative from the Research Unit for Musculoskeletal Function and Physiotherapy at the University of Southern Denmark with the overarching aim to implement current clinical guidelines for OA into clinical care. One part aims at patient education and neuromuscular exercise for patients with OA-like symptoms primarily from the hip or knee [27].

The sampling was purposive sampling. One author (AV) facilitated the contact to the supervising physiotherapist. The inclusion criterion was as follows: the participants had to have experiences with training and have an interest in using an exercise app. The exclusion criterion was as follows: people who were not able to understand Danish, as the app is in Danish.

Eight people initially agreed to participate; however, only 6 people with hip or knee arthritis participated in this study. Characteristics of the participants are shown in Table 1.

The app developer educated the participants in using the app before the test. They were giving oral and hands-on introduction to the app.

Roles and Relations With the Research Group

The research group was composed of both researchers with thorough experience with PD and researchers who had expertise in OA and who had been developing the app. The 3 researchers with expertise in PD were responsible for designing the different interviews and workshops, where the researcher with expertise in OA was in charge of recruitment of the participants, planning the testing of the app, and supporting the participants throughout the test.

The first author had team management skills and was responsible for the organizational and administrative procedures.

Data Collection

The data collection was divided into three processes (Figure 1).

Focus Group Interview

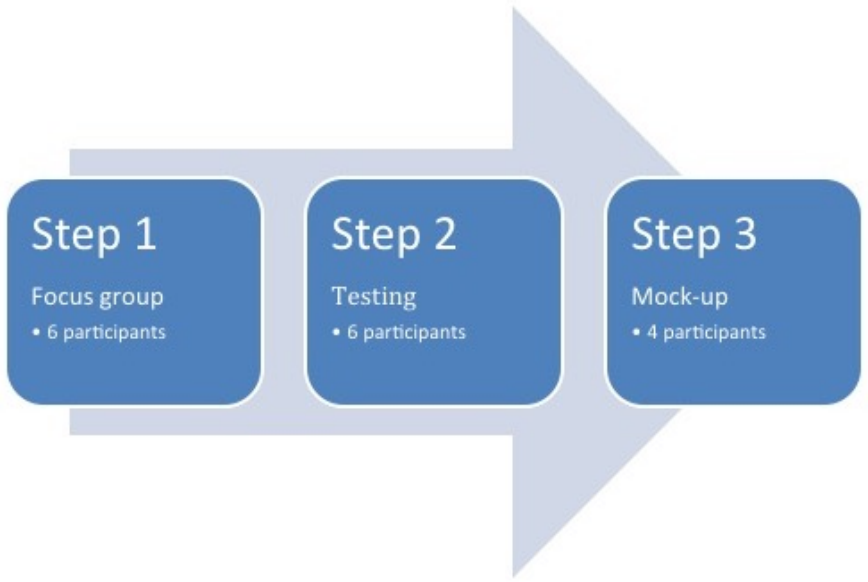
First, a focus group interview was conducted, as focus group discussions can mobilize associations , where the group dynamic contributes to the creation of narratives [28]. All 6 participants participated in the focus group.

Table 1. Characteristics of the participants.

Participant	Gender (female or male)	Ethnicity	Employment	Surgery (yes or no)	Daily exercise or activity
1	Female	Danish	Old age pensioner	No	Exercise on a daily basis
2	Male	Danish	Old age pensioner	Yes	Daily activity: walking the dog
3	Female	Danish	Employed	No	Daily activity: biking
4	Female	Danish	Employed—light duties	No	Daily activity: gardening
5	Female	Danish	—[a]	No	—
6	Male	Danish	Old age pensioner	No	Exercise on a daily basis

[a]— indicates missing data.

Figure 1. Three processes of data collection.



The focus group interview was conducted using open questions and some follow-up questions. This was to let the participants discuss freely on the topics, and allow for the possibility of asking follow-up questions, if the areas of interest for the research had not already been covered. An interview guide was compiled, and the two overall themes were on mental and physical barriers and motivational and social aspects of training at home for people with hip or knee OA. The first, second, and last authors conducted the focus group interview. It was the first author who was the primary moderator. The participants were asked how they would exercise when they were no longer part of a training class. Then, they were invited to write their reflections down, and then share them with the other group members. If needed, the moderator would ask a follow-up question and ensure that all group members shared their experiences.

At the end of the focus group, the participants were asked if they still wanted to try the training app. All participants agreed to try the app.

Testing of Existing App

Second, a test of the training app was initiated. The app is commercially available at App Store and Google Play Store. A user profile (email address and created password) is necessary to use the app. An underlying database holds individual user information such as exercise level, which is used for continuous individualization of exercise. The database also contains a pain and exercise diary, which is entered by the user. This allows the regeneration of the diary if it is lost from the user's device.

The testing was divided into a set of predefined tasks or functions (eg, download the app from marketplace, create a user profile, use existing exercise programs, and maintain an exercise and pain dairy). The test period was set to last 4 weeks, and personal reminders were sent weekly on which task to focus on. Participants could contact the author (AV) for support.

To capture the participants' experiences during the test, we used cultural probes as a mean of collecting data about their feelings and thoughts during the testing [29]. The probes are small packages that can include any sort of artifact. The package included a small notebook, a card with reflection questions for the participants, and a card with an invitation to take a photo of where and how they are trained.

Mock-Up Workshop

Finally, we met the participants for a mock-up workshop. Two of the participants did not participate. The purpose of the mock-up workshop was to gain knowledge of the participants' experiences arising from testing the app, and to gain ideas for further development of the app.

Mock-up is a creative method where the users and the researchers together transform the users' knowledge to solutions [30]. The starting point for the mock-up workshop was the cultural probes (from the test phase), which served as the opening to hear about the participants' experiences. Subsequently, the participants worked on idea generation. First, ideas were written on post-it notes; then, the best ideas were chosen and the participants created a mock-up model (Figures 2-4). The participants were all active in the *making* process, and they used both language and hands for expressing their thoughts and ideas. At the end of the workshop, the participants gave feedback to the mock-up model to ensure that the result reflected the participants' views on training.

The participants worked in a group where they had paper, colored pencils, pen, and felt-tip pens to use. The participants were introduced to the workshop for them to understand their role in designing the future app. The first, second, and last authors facilitated the workshop. It was the first author who was the primary facilitator.

Both the focus group and the workshop were audiotaped and transcribed. The focus group lasted 90 min, and the workshop lasted 120 min.

Figure 2. Suggestions for features in the app. The features are: introduction, music, camera, TV, alarm, award, age, personal trainer, and overall assessment.

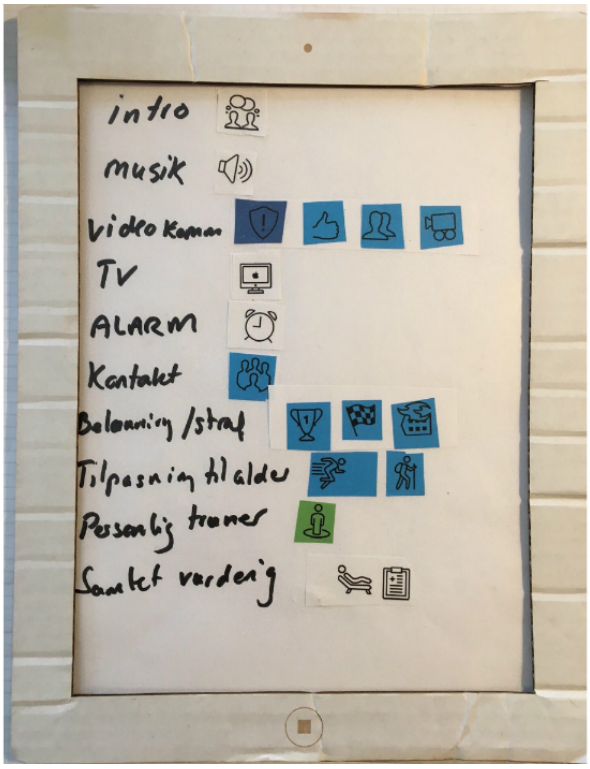


Figure 3. The different features reflecting the participants' ideas about interaction.

Figure 4. The different features reflecting the participants' ideas about motivation and reminders.

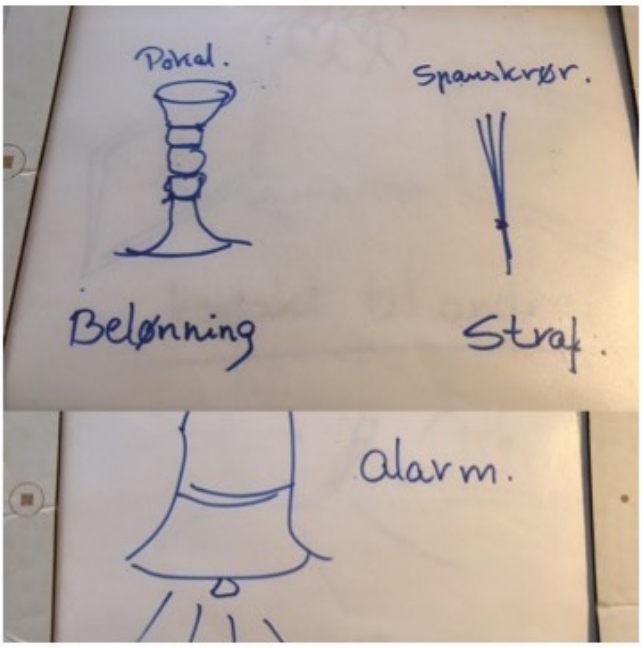


Table 2. Process of analysis: examples from the analysis.

Step 1: superior themes extracted after the first open reading	Step 2: From themes to codes. Identifying meaningful units. The meaningful units are coded based on the superior themes		Step 3: From codes to meaning. The meaningful units are sorted into groups
	Quotations	Code	
Competitive	*But when you can see that someone did something, then you can think to yourself that maybe I should try and do that too. Not that it is a competition but a way to challenge myself. And then I also write it down (P2)*	*Motivation*	*Competition as motivation*
Prefer being together	*I am that kind of type that feel most comfortable doing it together with someone else (P4)*	*Barrier*	*Barriers*

Data Analysis

The data from the focus group interview and the workshop were analyzed separately and results will be presented accordingly. The data from the focus group interview were used to identify the mental and physical barriers and motivational and social aspects of training at home for people with hip or knee OA. The data from the test and the mock-up workshop were used to get ideas for the further development of the app. The data analysis was inspired by Malterud's systematic text condensation [31] and organized according to the steps taken in the analysis, as shown in Table 2.

First, we captured an overall impression of the data and extracted a preliminary set of main themes. Second, data were divided into meaningful topics, which were relevant to the study question. Then, the meaningful topics were condensed and coded. Finally, the findings were synthesized, involving a shift from condensation to descriptions and categories. The codes were developed based on the preliminary themes identified in the first step and the theoretical framework.

To enhance validation, the first and third authors worked on the analysis together. The analysis was performed with the transcripts printed. They discussed their overall impression of the data and then they highlighted the meaningful topics with a marker, and the codes were discussed between the 2 authors. The first author wrote down the analysis. They discussed the analysis with the other authors afterward. The findings were then discussed in relation to relevant literature and theory on motivation.

Ethical Considerations

The participants were informed both orally and in writing about the study, and were included after providing their informed consent in compliance with the Helsinki Declaration [32].

The study was submitted to the Scientific Ethics Committee. The committee decided that approval from an ethics committee was unnecessary according to national legislation in Denmark. The Danish Data Protection Agency registered the study (2008-58-0035).

Results

Results from the focus group interview revealed 3 categories reflecting the participants' experiences with training and their motivation and the barriers for training.

Results From the Focus Group Interview

Competition as Motivation

Most of the participants could report that competition played a central role in their motivation for exercise. They expressed that competition could be a motivational factor—not in the sense that they were competing against each other, but more a competition with yourself. It also became clear that it was important to track exercise to be able to see progress. One participant stated:

> *When you can see that someone did something, then you think to yourself, that maybe I should try and do that too. Not that it is a competition but a way to challenge myself. And then I also write it down.* [P2]

All participants were aware that each person had different challenges, and for that reason, it did not make sense to compete against each other; they were just more motivated by seeing each other's progress and to see if they could do the same things as their peers.

Training Together

The majority of the participants reported that avoiding surgery was a strong motivation. One participant stated:

> *I found out that by doing my training, I don't need surgery on my left knee, this keeps me going, that I don't have to go through it once again.* [P2]

Most of the participants believed that training together was the main motivational factor. One of the participants stated that although she had been in pain for years and felt sorry for herself when the doctor told her that she did not need a new knee, she did not do training. It was not until she joined a training team. She stated:

> *I didn't do my training, but when I joined GLA:D then I did, and it went really well. And it has given me great results […].* [P3]

All the 4 female participants reported that training in a group motivated them. One participant reported that she had a difficult time doing her exercises besides the supervised exercise because she needed someone to remind her . She stated:

> *But getting it done besides at the class…No it would be helpful if someone tapped my shoulder and said hey you need to do it now.* [P5]

All participants agreed that some kind of a reminder would be helpful; one of the participants had put elastic bands around the coffee table as a reminder to do exercise when watching television. One participant stated:

> *It actually means that when I sit down, I think well I can just as well do the exercises while I watch the news.* [P1]

Barriers

One of the participants reported she felt that the training helped her but could not keep the motivation when she stopped attending the class . She stated:

> *Then I stopped, and I am back into the routines of my daily life […] I have no backbone.* [P3]

She was not the only one who felt motivation as a barrier to do the exercises on her own.

One of the men reported that he trained by himself and his motivational factor was his daily walks with his dog . He stated:

> *I'm motivated by my dog and fresh air—training enhances my self-esteem.* [P6]

One of the women who had a hard time motivating herself at home also made a point that when she did not feel pain, she did not train, but if the pain came back, she would start all over . She stated:

> *I am aware of it, and if it starts to hurt again, I will do it.* [P3]

Another aspect that made it difficult for the participants to perform the exercise alone at home was the doubt whether they were doing the exercises correctly. The participants reported this as a reason that they stopped training. One participant stated:

> *I need to have some input from a professional, because it makes me aware of what I am doing and how I am moving.* [P4]

Results from the Workshop

The analysis of the data from the workshop reflected how the participants had experienced using the app, their thoughts on motivation and barriers, and their ideas about creating the paper prototype.

It is presented in the following categories.

User Experiences

The participants shared their experiences using the app. Participants experienced different types of technical difficulties when downloading the app, or when trying to load a video—for example, the screen "freezing." Despite these initial problems, they had all figured out how to use the app. One of the participants had sent an email to the research team, but before he had replied to her, she had found out by herself.

The participants had figured out by themselves how to use the app, but they would have liked an introduction to the app or more precisely to the exercises. One participant stated:

> *It looks different when you look at yourself, when you make an exercise than if you look at someone else. And to have someone correcting you. We know how important it is to perform the exercise correctly because we can make damage if we don't do it the right way.* [P4]

They all stated that a physical meeting before starting using the app is necessary, but they also discussed that the app could be the follow-up offer after their participation at the GLA:D class ended. One participant stated:

> *I think that if we had gotten the app earlier, and we could have used it as a continuation of the class, then it had been easier and I would have continued with the training, and then we could also have had the physical therapist words in the back of our minds.* [P3]

It was also reported that instead of using the app, they had used the exercise instructions on paper sheets that was handed out at the class because it was familiar and they felt confident in the familiarity of the exercises.

They all agreed that they missed the class, and they would have liked something after the classes stopped. One of the participants had written the following in her notebook (handed out as part of the cultural probe):

> *You wrote the question: "do you miss being in the GLA:D class?" And I wrote "The class made me feel obligated to go. Now it is easier to find excuses not to do the training. And what I miss the most is the professional guidance when we were together with the physical therapists.* [P3, note in notebook, cultural probe]

Motivation

One of the participants saw that the app could help her in different ways, both to get the actual exercises done and also to register different activities such as biking, running, and gardening. She needed to lose weight and saw it as a way to keep track of her activity level. However, she pinpointed how the social aspect of being in a class was important for her training and thereby her rehabilitation. She stated:

> *The motivation is there, but then again—it is in an app, and I am better in the class. And there I remember that I'm not the only one who feels that way.* [P3]

The participants reported that the app was easy to use, which they found motivating. They also found that the exercises were good. However, they all agreed that they needed something more that could make them stick to the training. One of the participants had been involved in a project where she should write in a diary how she felt and experienced the trainingand rehabilitation. Every month, she received a message if she had not done it. She was motivated by that kind of monitoring.

The participants also discussed that although they ought to take responsibility for their own health, it had a positive effect if they knew someone else was "watching." She stated:

> *It is a great help that someone is saying "hello" (she knocks the table to underline the meaning).* [P1]

Ideas for a Prototype

The participants made a mock-up of how the app should look to meet their needs and thereby addressing their barriers/enablers that were identified in the focus group interview. The following section is a description of their ideas, and some of their drawings are selected to show how they worked with designing the paper prototype.

Figure 2 shows the suggestions for features in the app. The features are: introduction, music, camera, television, alarm, award, age, personal trainer, and overall assessment.

The ideas were categorized into the following: interactive, motivation, and user experiences.

Interactive

The participants requested an introduction to the app. It could be at a physical meeting with a physiotherapist or via a feature in the app. They suggested that it could be a visual guidance to the different exercises supplemented with text. In addition, they asked for the possibility to have video calls where they could get advice and have their exercises corrected.

It was also important to the participants that they could have contact with other patients, as it was highlighted during the first focus group interview that the participants were motivated by training together with others and by seeing each other's progress and to see if they could do the same things as their "class mates" or peers. This could, for instance, be a chat forum where they could share their experiences with training.

Motivational features

The participants had several suggestions for different motivational features. For instance, they could either get rewarded or be given a "penalty." In addition, they suggested that the app should have an alarm where they were notified that it was time to exercise. This could also meet the needs of the majority of the participants who in the focus group interview reported that they needed someone to remind them.

Individualized features

The participants wished for more individualized features to make the user experience better. For instance, they would like to be able to choose music. They would also like to have the exercises shown on their television.

Discussion

Principal Findings

OA is a chronic and degenerative joint disease, which causes pain and decreased physical function. Exercise has proven effective in diminishing pain and to postpone the need for surgical intervention. However, the lack of continuous adherence to exercise remains a challenge for people with OA as well as patients with other chronic diseases and obesity.

The results from this study revealed that all the participants wanted to do their training and knew that it was the best way to avoid operation and to minimize their pain. Hence, half of the participants described being motivated on one hand to avoid operation and unnecessary pain, but others found it hard to motivate themselves to do the exercises on their own. They missed the observation, comments, and encouragement by the supervising physiotherapist as well as their peers. They described motivation in the exercise community as the feedback from the instructor, and they were also depending on "the tap on the shoulder" to get going. Nevertheless, some of the patients created their own tailored activities and included these in their

daily life, which became a driving factor for sustainable motivation.

The participants made a mock-up of how the app should be redesigned to meet their needs and thereby addressing their barriers or enablers that existed.

Motivation—Internal and External Facilitators and Barriers

Motivation was a major issue for the participants. Motivation is the element within the individual, which evokes and maintains certain behaviors [33]. Motivation can shed light on the reasons for someone to act in a certain way [34]. Different things motivated the participants in our study. Among other things that motivated them, they highlighted competition (both with themselves and each other), being social, avoiding pain, avoiding surgery, and being able to walk the dog. These different things can be characterized as *internal* and *external* motivation [33]. External motivation is created from outside the person, and cannot be controlled by an individual. Internal motivation is created within the individual, and the behavior occurs, as it is satisfying for the person. An important aspect of the internal motivation is a feeling of being *capable*. This can be enhanced by experiences of success when training. In this study, all participants highlighted the importance of contact with the therapist to be corrected in the way of doing the exercises and at the same time getting reassurance from the expert, and it was also reflected in the redesigning of the app. This may have to do with the participants not feeling competent while doing the exercises alone. Feeling competent may increase the participants' motivation to training. Similarly, feeling unsuccessful can weaken their motivation. This can be linked to the concept of self-efficacy. Self-efficacy refers to the belief in one's ability to successfully perform a particular behavior [35]. Bandura's theories originated in behaviorism but took a more humanistic approach, with a focus on the social, biological, and cognitive aspects of learning. Bandura's social learning theory involves the concept of self-efficacy. Bandura states that self-efficacy beliefs influence the way that people think, feel, and act [35]. For patients to positively engage exercise behavior, they must have confidence in performing the specific behavior. Patients with high self-efficacy are likely to make a greater effort than patients with low self-efficacy. According to Bandura, perceived self-efficacy plays a key role in adapting to the new behavior. Self-efficacy beliefs are built either through one's own experiences (mastery experiences), other experiences (vicarious experiences), support from people in one's environment (verbal persuasion), or through emotional experiences (physiological and affective state of mind).

According to Bandura, self-efficacy beliefs should incorporate the level of specific knowledge pertaining to the actions involved in training as well as confidence in one's ability to carry out the specific activities [36]. Previous experiences, both positive and negative, as well as a lack of experiences, have an impact on patients' perceptions of efficacy. Psychosocial mood also has an influence on the experiences. A positive attitude toward training, a good experience of training, and a positive state of mind also positively affect training experiences [36].

Vicarious experiences, as well as social and verbal persuasion, from family, peers, and therapists contribute to self-efficacy. This underlines the importance of both having good experiences with training, as well as the possibility of training together, and the feedback from a therapist [35,36]. It was also highlighted in the redesigning of the app where the participants requested a chat room, where they could connect with peers.

Petursdottir et al [37] identified facilitators and barriers, which will influence a person with OA exercise behavior. The internal facilitators and barriers are both personal experiences such as effect on pain, finding suitable exercise, and the benefits of exercising. This was also found in this study—the 2 men who participated were both motivated by internal facilitators, whereas the 4 women were all motivated by the external facilitators. The external facilitators described by Petursdottir et al were the physiotherapist' professional care, training partners, and the availability of exercise classes.

Pertursdottir et al found the effect on pain to be the most significant factor. In this study, the participants also emphasized this as an important facilitator; however, the participants could also report that when the pain was gone, it was hard for them to continue training, although they knew that there was a risk that the pain would return.

The participants also underlined that it was difficult to keep up with the training after stopping the supervised class because they got out of the training routine and fell into their daily routine, and the participants also expected that the technology should be able to help them to keep up with a daily training routine both by having a reminder function and the possibility to connect with other peers. This can be explained by the literature, where other studies found that an important factor for the patients to do the exercises is being part of an organized training activity [37,38].

Exercise Community

The majority of the participants stressed that the company of others motivated them and that the training was conducted in a class and with guidance from a therapist. This is found in other studies as well, where social support is highlighted as a significant motivational factor [37,39,40].The support can be from the person's family and friends, as well as health care professionals, or from a training class.

The majority of the participants were depending on the social support, and although they all agreed that they had the main responsibility themselves, they were motivated if someone kept them on a short leash. This can be explained with a behavioristic view of the changing behavior, where changing behavior occurs with either a reward or penalty. As well with the concept of self-efficacy, where learning from others, play an important role in building your self-efficacy [35]. This is essential to consider when designing a training app.

Technology as a Way to Overcome the Barriers

Other studies have shown how the use of interactive technology can motivate participants. Thorup et al found that a pedometer could offer independence from standardized rehabilitation as it could individualize the walking activity based on the patient's

choice. The pedometer delivered feedback on walking activity, which led to an increased competence for the patients to achieve their goals for steps [41]. This can be linked to the motivational theory, where feeling competent while doing the exercises can support the participants' motivation [33].

Thorup et al also found that the pedometer supported relatedness with others. The health professionals' surveillance of patients' steps made the patients feel observed and supported [41]. This is interesting in regard to our findings, where the participants were depending on the social and interactive aspects of training, which was reflected in their demands for the redesigning of the app. This is a way in which the app can be further developed, thereby overcoming the barriers identified.

There is no evidence published on whether technologies can influence patients with OA training and their outcomes. There are some studies showing that consistent contact via phone can improve the clinical status of patients with knee OA [7], whereas a randomized controlled trial revealed that monthly phone contact aimed at promoting self-care for patients with knee OA could be associated with improvements in joint pain and physical function [42]. This is consistent with the findings of our study, where the personal contact is an external motivation for the participants. This supports the participants' ideas for interactive elements in the app, and also their experiences using the app, where they found it easy to use, and they also found that the exercises were good, but it was not enough for them to use it. The participants had ideas to support the experience of "going to class" and to get feedback from a therapist.

The idea of being *controlled* and receiving reminders, as well as rewards or penalty, corresponds well with the use of automated interactive technology.

One of the most usual apps is a reminder for patients to take their medications with the use of text messaging [43]. In addition, there is the potential that redesigning the app with a reminder function can encourage the participants' training behavior. As shown in the study of the pedometer, the surveillance of patients' steps made the patients feel observed and supported and thereby encouraged them to do the exercises [41,44]. One of the participants explained how she had customized her coffee table for exercise with the use of elastic bands, so she could do her training while watching the news. It shows how small adjustments to the daily routines can change behavior. This can be captured in a training app, where regular messages are customized for the individual to enhance training.

Strengths and Limitations

The limitation of this study is that it was a small-scale study; however, most qualitative studies are typically small-scale. Therefore, despite the small sample size, the aim of this study, as other qualitative studies, was to provide in-depth exploration of the phenomenon under investigation. Therefore, the intention of this study was to understand and explain the mental and physical barriers and motivational and social aspects of training at home for people with hip or knee arthritis and to enhance and test an exercise app for use at home by patients with hip or knee OA.

However, it has been taken into account that only 6 patients with OA from a specific medical center or community center were included, and a wider representative sample would have provided more in-depth information. Future studies may have to consider this.

We have provided rich descriptions of both the mental and physical barriers and motivational and social aspects of training at home for people with OA, as well as rich and visual descriptions of their experiences with an exercise app and also the suggestions for further development of the app.

This will hopefully allow the readers to judge whether the work is potentially transferable to their own contexts. The results cannot claim statistical generalizability, but analytical generalization [45], which emerges by means of the dialectic between theory and practice.

The analysis was conducted together with coresearchers to increase the reliability, and we presented the analysis process in a table to make the analysis transparent. Quotations from the focus group interview were used to link to the participants' original statements to warrant validity.

Conclusions

The conclusion of the study is that the long-term continuation of exercising for patients with OA could be improved with the use of a technology tailored to users' needs, including motivational and other behavioral factors. The study highlighted that the continuation of rehabilitation is easiest in the short term when the benefit for the patient is visible and rewarding. In the long term, it takes more motivation to continue—motivation was often facilitated by a physical meeting with the therapist. We need to find new ways of connecting the therapist and peers to the patient's daily life, and health technology as a tailored app seems to hold promises.

The participants were informed both orally and in writing about the study, and were included after providing their informed consent in compliance with the Helsinki Declaration [32].

The study was submitted to the Scientific Ethics Committee. The committee decided that approval from an ethics committee was unnecessary according to national legislation in Denmark. The Danish Data Protection Agency registered the study (2008-58-0035).

Acknowledgments

The authors wish to thank the study participants and the rehabilitation department of Slagelse municipality. No funding was obtained for this study.

Authors' Contributions

DBD was involved in the study design and conducted the focus group interviews and workshops. She analyzed the data and was the primary writer of the manuscript. AV was involved in the study design and recruitment of participants and administrative personnel in Slagelse municipality. He assisted in manuscript preparation. GV transcribed the data and was involved in the analysis and manuscript preparation. MJR was involved in the analysis of data and manuscript preparation. JC was involved in the study design and conducted the focus group interviews and workshops. She was involved in the analysis of data and manuscript preparation.

Conflicts of Interest

AV is the owner and developer of the app, Ther-ex. The app is available without costs at App Store and Google Play Store and contains no advertisements.

References

1. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med 2000 Oct 17;133(8):635-646. [Medline: 11033593]
2. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010 Aug;26(3):355-369 [FREE Full text] [doi: 10.1016/j.cger.2010.03.001] [Medline: 20699159]
3. Bitton R. The economic burden of osteoarthritis. Am J Manag Care 2009 Sep;15(8 Suppl):S230-S235 [FREE Full text] [Medline: 19817509]
4. Leigh JP, Seavey W, Leistikow B. Estimating the costs of job related arthritis. J Rheumatol 2001 Jul;28(7):1647-1654. [Medline: 11469474]
5. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010 Apr;18(4):476-499 [FREE Full text] [doi: 10.1016/j.joca.2010.01.013] [Medline: 20170770]
6. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, European League Against Rheumatism (EULAR). EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013 Jul;72(7):1125-1135 [FREE Full text] [doi: 10.1136/annrheumdis-2012-202745] [Medline: 23595142]
7. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008 Feb;16(2):137-162 [FREE Full text] [doi: 10.1016/j.joca.2007.12.013] [Medline: 18279766]
8. Medizinische Universitat Wien. 2012. Standards of Care for OA and RA URL: http://www.eumusc.net/myUploadData/files/Standards%20Of%20Care%20Full%20report.pdf [accessed 2017-12-28] [WebCite Cache ID 6w2iEkqoZ]
9. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res 2012 Mar 27;64(4):465-474. [doi: 10.1002/acr.21596]
10. Skou S, Simonsen M, Odgaard A, Roos E. Predictors of long-term effect from educationexercise in patients with kneehip pain. Dan Med J 2014 Jul;61(7):A4867. [Medline: 25123117]
11. Thorstensson CA, Garellick G, Rystedt H, Dahlberg LE. Better management of patients with osteoarthritis: development and nationwide implementation of an evidence-based supported osteoarthritis self-management programme. Musculoskeletal Care 2015 Jun;13(2):67-75. [doi: 10.1002/msc.1085] [Medline: 25345913]
12. Ganz D, Chang J, Roth C, Guan M, Kamberg C, Niu F. Quality of osteoarthritis care for community-dwelling older adults. Arthritis Rheum 2006;55(2):241-247. [doi: 10.1002/art.21844] [Medline: 16583414]
13. Alami S, Boutron I, Desjeux D, Hirschhorn M, Meric G, Rannou F. Patients' and practitioners' views of knee osteoarthritis and its management: a qualitative interview study. PLoS One 2011;6(5). [doi: 10.1371/journal.pone.0019634] [Medline: 21573185]
14. Holden M, Nicholls E, Young J, Hay E, Foster N. UK-based physical therapists' attitudes and beliefs regarding exercise and knee osteoarthritis: findings from a mixed-methods study. Arthritis Rheum 2009;61(11):1511-1521. [doi: 10.1002/art.24829] [Medline: 19877105]
15. Ettinger WJ, Burns R, Messier S, Applegate W, Rejeski W, Morgan T. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). J Am Med Assoc 1997;277(1):25-31. [Medline: 8980206]
16. van Baar M, Dekker J, Oostendorp R, Bijl D, Voorn T, Bijlsma J. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. Ann Rheum Dis 2001;60(12):1123-1130. [Medline: 11709454]
17. Bennell K, Hinman R, Wrigley T, Creaby M, Hodges C. Exercise and osteoarthritis: cause and effects. Compr Physiol 2011 Oct;1(4):1943-2008. [doi: 10.1002/cphy.c100057] [Medline: 23733694]
18. Bennell K, Dobson F, Hinman R. Exercise in osteoarthritis: moving from prescription to adherence. Best Pract Res Clin Rheumatol 2014;28(1):93-117. [doi: 10.1016/j.berh.2014.01.009] [Medline: 24792947]

19. Kloek C, Bossen D, Veenhof C, van Dongen J, Dekker J, de Bakker D. Effectiveness and cost-effectiveness of a blended exercise intervention for patients with hip and/or knee osteoarthritistudy protocol of a randomized controlled trial. BMC musculoskeletal disorders 2014;15(1):269. [Medline: 25103686]
20. Bossen D, Veenhof C, Dekker J, de Bakker D. The usability and preliminary effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis. BMC Med Inform Decis Mak 2013;13(1):61. [doi: 10.1186/1472-6947-13-61]
21. Bossen D, Veenhof C, Dekker J, de Bakker D. The effectiveness of self-guided web-based physical activity interventions among patients with a chronic disease: a systematic review. J Phys Act Health 2014;11(3):665-677. [Medline: 23493018]
22. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
23. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
24. Simonsen J, Robertson T. Routledge international handbook of participatory design. New York: Routledge; 2012.
25. Gadamer H. Truth and method. London: Bloomsbury; 2013.
26. Kensing F. Methods and practices in participatory design. Copenhagen: ITU Press Copenhagen; 2003.
27. University of Southern Denmark. Gla:D. 2012. Good Life with osteoArthritis in Denmark (GLA:D) URL: https://www.glaid.dk/english.html [accessed 2017-12-21] [WebCite Cache ID 6vs2Ilw7q]
28. Barbour R. Doing focus groups. London: Sage; 2018.
29. Gaver B, Dunne T, Pacenti E. Design: cultural probes. Interactions 1999;6(1):21-29. [doi: 10.1145/291224.291235]
30. Brandt E, Binder T, Sanders B, Simonsen J, Robertson T. Tools and Techniques - ways to engage telling, making and enacting. In: Simonsen J, Robertson T, editors. Routledge International Handbook of Participatory Design. New York: Routledge; 2013.
31. Malterud K. Qualitative methods in medical research. 2nd edition. Oslo, Norway: Universitetsforlaget; 2003.
32. World Medical Association. 2008. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects URL: http://www.wma.net/en/30publications/10policies/b3/ [accessed 2017-12-21] [WebCite Cache ID 6vs2l7NAm]
33. Hein H. Motivation: motivationsteori og praktisk anvendelse. Copenhagen: Hans Reitzels Forlag; 2009.
34. Beyer N, Lund H, Klinge K. Træning i forebyggelse, behandling og rehabilitering - Excercise in prevention, treatment and rehabilitation. Denmark: Munksgaard; 2010.
35. Bandura A. Self-efficacy: The exercise of control. New York: W.H. Freeman and Company; 1997.
36. Bandura A. Regulation of cognitive processes through perceived self-efficacy. Dev Psychol 1989;25(5):729.
37. Petursdottir U, Arnadottir S, Halldorsdottir S. Facilitators and barriers to exercising among people with osteoarthritis: a phenomenological study. Phys Ther 2010 Jul;90(7):1014-1025. [doi: 10.2522/ptj.20090217] [Medline: 20466741]
38. Hendry M, Williams NH, Markland D, Wilkinson C, Maddison P. Why should we exercise when our knees hurt? A qualitative study of primary care patients with osteoarthritis of the knee. Fam Pract 2006 Oct;23(5):558-567. [doi: 10.1093/fampra/cml022] [Medline: 16731544]
39. Damush TM, Perkins SM, Mikesky AE, Roberts M, O'Dea J. Motivational factors influencing older adults diagnosed with knee osteoarthritis to join and maintain an exercise program. J Aging Phys Act 2005 Jan;13(1):45-60. [Medline: 15677835]
40. Holden MA, Nicholls EE, Young J, Hay EM, Foster NE. Role of exercise for knee pain: what do older adults in the community think? Arthritis Care Res (Hoboken) 2012 Oct;64(10):1554-1564 [FREE Full text] [doi: 10.1002/acr.21700] [Medline: 22511582]
41. Thorup C, Grønkjær M, Spindler H, Andreasen J, Hansen J, Dinesen B. Pedometer use as motivation for physical activity in cardiac tele-rehabilitation. Int J Integr Care 2015;15(7). [doi: 10.5334/ijic.2288]
42. Weinberger M, Tierney WM, Booher P, Katz BP. Can the provision of information to patients with osteoarthritis improve functional status? A randomized, controlled trial. Arthritis Rheum 1989 Dec;32(12):1577-1583. [Medline: 2688659]
43. Jones D, Skrepnik N, Toselli RM, Leroy B. Incorporating novel mobile health technologies into management of knee osteoarthritis in patients treated with intra-articular hyaluronic acid: rationale and protocol of a randomized controlled trial. JMIR Res Protoc 2016;5(3):e164. [Medline: 27506148]
44. Thorup C, Grønkjær M, Spindler H, Andreasen J, Hansen J, Dinesen B. Pedometer use and self-determined motivation for walking in a cardiac telerehabilitation program: a qualitative study. BMC Sports Science, Medicine and Rehabilitation 2016;8:24. [doi: 10.1186/s13102-016-0048-7]
45. Brinkmann S. Qualitative Interviewing. New York: Oxford University Press; 2013.

Abbreviations

GLA:D: Good Life with osteoArthritis in Denmark
OA: osteoarthritis
PD: participatory design
Ther-ex: Therapeutical exercise

Edited by G Eysenbach; submitted 22.03.17; peer-reviewed by F Vasefi, N Bragazzi, H Devan; comments to author 09.08.17; revised version received 22.08.17; accepted 16.11.17; published 11.01.18.

Please cite as:
Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J
Usage of an Exercise App in the Care for People With Osteoarthritis: User-Driven Exploratory Study
JMIR Mhealth Uhealth 2018;6(1):e11
URL: https://mhealth.jmir.org/2018/1/e11/
doi:10.2196/mhealth.7734
PMID:29326092

Original Paper

Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability

Gitanjali Das[1,2], BS; Cynthia Cheung[3], MPH, MA; Camille Nebeker[1,3], EdD, MS; Matthew Bietz[3,4], PhD; Cinnamon Bloss[1,3,5], PhD

[1]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, United States
[2]Division of Biological Sciences, University of California San Diego, La Jolla, CA, United States
[3]Qualcomm Institute, California Institute for Telecommunications and Technology, University of California San Diego, La Jolla, CA, United States
[4]Donald Bren School of Information and Computer Sciences, Department of Informatics, University of California Irvine, Irvine, CA, United States
[5]Department of Psychiatry, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Cinnamon Bloss, PhD
Department of Psychiatry
University of California San Diego
9500 Gilman Drive
Atkinson Hall
La Jolla, CA, 92093-0811
United States
Phone: 1 858 534 9595
Fax: 1 858 534 9404
Email: cbloss@ucsd.edu

Abstract

Background: Due to the growing availability of consumer information, the protection of personal data is of increasing concern.

Objective: We assessed readability metrics of privacy policies for apps that are either available to or targeted toward youth to inform strategies to educate and protect youth from unintentional sharing of personal data.

Methods: We reviewed the 1200 highest ranked apps from the Apple and Google Play Stores and systematically selected apps geared toward youth. After applying exclusion criteria, 99 highly ranked apps geared toward minors remained, 64 of which had a privacy policy. We obtained and analyzed these privacy policies using reading grade level (RGL) as a metric. Policies were further compared as a function of app category (free vs paid; entertainment vs social networking vs utility).

Results: Analysis of privacy policies for these 64 apps revealed an average RGL of 12.78, which is well above the average reading level (8.0) of adults in the United States. There was also a small but statistically significant difference in word count as a function of app category (entertainment: 2546 words, social networking: 3493 words, and utility: 1038 words; P=.02).

Conclusions: Although users must agree to privacy policies to access digital tools and products, readability analyses suggest that these agreements are not comprehensible to most adults, let alone youth. We propose that stakeholders, including pediatricians and other health care professionals, play a role in educating youth and their guardians about the use of Web-based services and potential privacy risks, including the unintentional sharing of personal data.

(JMIR Mhealth Uhealth 2018;6(1):e3) doi:10.2196/mhealth.7626

KEYWORDS

privacy; comprehension; mobile applications; adolescent

Introduction

Both Apple and Android have recently surpassed 1.5 million apps available on their respective markets [1]. Most of these apps collect user statistics and are able to make use of the built-in sensors on one's mobile phone to track movement, location, and other personal behavior and activity [2]. Although the use of built-in sensors may simplify the user interface and improve user experience, it can also allow app developers and third parties to gather potentially sensitive information about the consumer [2]. Due to the growing availability of consumer information, protection of personal data is of increasing concern.

Privacy policies should inform users of the risks of the product they are about to use. Whereas most users may not read the privacy policy, if they have concerns about their privacy while using an app, they should be able to refer back to the policy to understand how their information is being collected or used. Although the Federal Trade Commission (FTC) recommends that mobile apps make privacy statements available to app users [3], not all apps have privacy policies. Furthermore, there are no clear standards regarding the accessibility of privacy statements for the average consumer, so privacy policies are often lengthy and difficult to read and comprehend [4]. In fact, an analysis of the privacy policies of mobile health apps conducted in 2015 found that most mobile health apps did not have privacy policies. Of the privacy policies that did exist, two-thirds of them did not focus solely on the app itself but instead addressed several apps or services offered by the developer. The available policies also did not make privacy practices transparent to the readers and had a high reading grade level (RGL) [4]. This presents a unique set of challenges when considering apps targeted toward minors.

Two existing regulations have attempted to address these issues: the FTC's Children's Online Privacy Protection Act (COPPA) and the California Online Privacy Protection Act (CalOPPA). The COPPA took effect in 2000 and created stipulations for the collection, usage, and sharing of information from children under 13 years by Web-based services. In 2013, COPPA rules were updated to address the privacy threats associated with "big data" and the ability for mobile apps and websites to collect highly granular information from consumers such as geolocation, relationships with friends, and different behaviors and preferences. The new COPPA guidelines also addressed parental concerns about websites collecting information about location, friends and contacts, and tracking software associated with mobile apps [5]. Similarly, the CalOPPA imposed regulations on apps available to California residents, requiring them to have a privacy statement informing consumers how their information is collected and shared [6]. CalOPPA also requires privacy statements to include a list of personally identifiable information being collected and a list of third parties with whom information is shared [6]. Unfortunately, it is still often unclear how third parties are collecting information that is entered into the app [7]. This calls into question the effectiveness of such a policy if users are not aware that apps are collecting their information.

The unnoticed involvement of third parties is of particular concern when considering apps targeted toward minors. Although the COPPA legally restricts the ways in which information from minors younger than 13 years can be collected and used, language in the COPPA excludes teenagers from 13 to 18 years of age from these same protections. Although the responsibility of monitoring a child's Web safety has traditionally fallen on the child's parents [8], the teenage years are a time when parents tend to have less direct oversight of Web-based activities. Teens who use mobile apps and websites are less likely to involve their parents when interfacing with and providing information to Web-based services [9] and may not be fully aware of how their information is collected and used. An open question, then, is the extent to which parents are able to adequately understand and advise on the privacy implications of their children's Web-based activities.

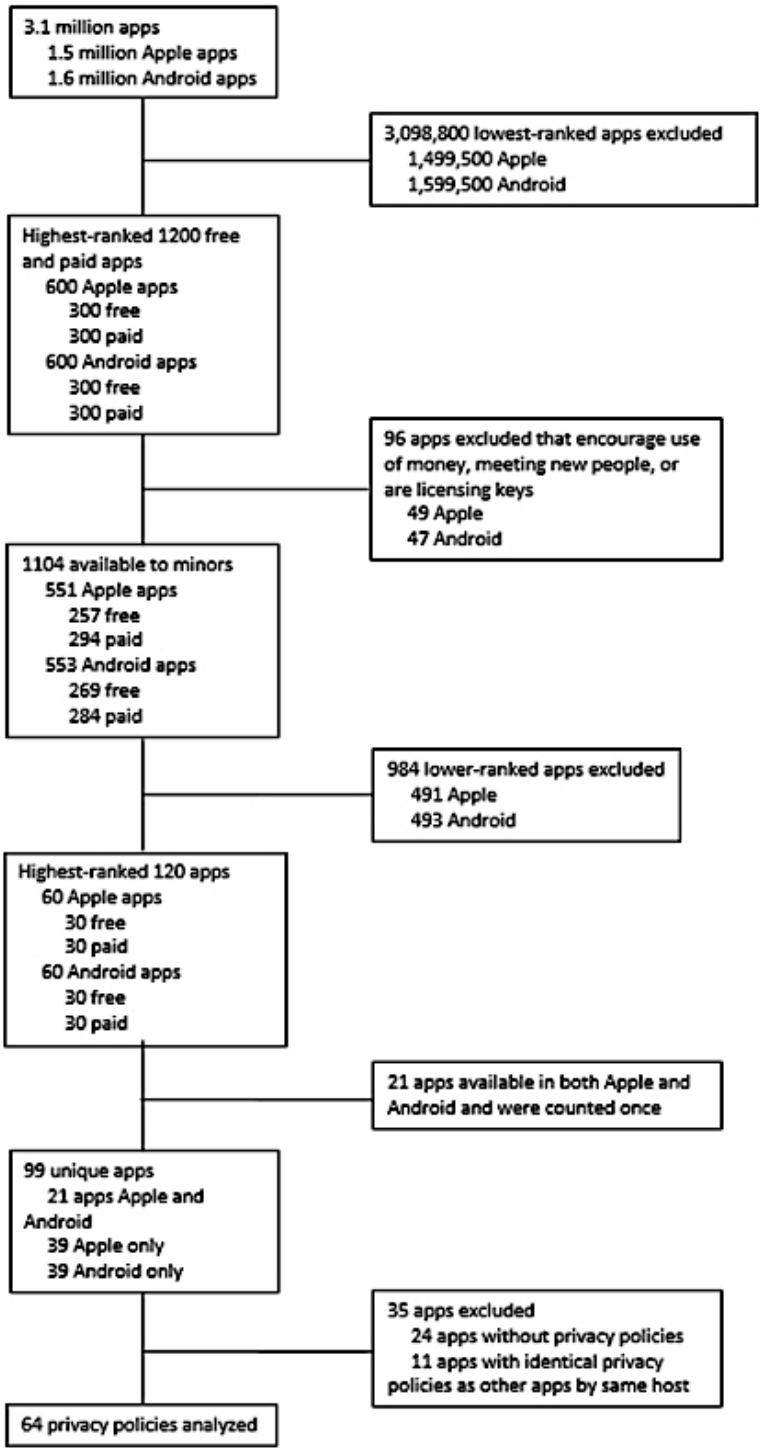
Internet safety has become a public health issue that concerns health care providers. The American Academy of Pediatrics (AAP) encourages parents to open a dialogue with their children about Web safety [10]. However, the lack of parental involvement in Web-based activity potentially leaves teens in a vulnerable situation regarding personal privacy and Web-based behaviors. For example, location tracking is a known safety concern particularly for teenage girls [9], making it important for teens to be aware of location-tracking features on the apps they download. Additionally, the increasing prevalence of social networking features in popular apps can expose youth to cyberbullying or unsuitable material, which can lead to long-term mental health consequences [11]. Although some research has shown that teenagers will take steps to protect their privacy by avoiding apps or disabling features that track their movements or usage [9], it is unclear whether the majority of teens are actually aware of the need to take such measures. Given that an estimated one in 3 Internet users is younger than 18 years [12], the implications of this issue are considerable. A 2012 analysis of app permissions and risk signals concluded that popular apps require more permissions for greater functionality, yet there are no reliable "risk signals" that alert users to the privacy risks associated with the app [13]. Privacy policies, such as an informed consent document, should be written in a way for users to understand their privacy risk when using an app. This study was designed to evaluate the readability of privacy policies for apps that are available to and targeted toward youth. Our goal was to inform strategies to educate and protect youth from unintentional sharing of personal data. The overarching privacy principles state that patients must be able to easily find and read the privacy policy of their health technology, and they have the right to refuse participation. The readability statistics collected in this study are compared with the Patient Privacy Rights' Trust Framework (PPR TF) principle #1 criteria on ability to find and understand privacy policies, which recommends an RGL of 12.0 or lower and a Flesch reading ease of 45.0 [14].

Methods

App Selection Process

Figure 1 outlines the app selection process used. The Apple App Store and the Google Play Store have a combined total of over 3 million apps available for download on mobile devices [15,16]. Each store ranks their apps according to their respective ranking formulas, which take into account app ratings, reviews, and number of downloads. We identified and analyzed the highest-ranked 300 free and 300 paid apps in the Apple App Store and the highest-ranked 300 free and 300 paid apps in the Google Play Store, for a total of 1200 apps, which were reviewed manually.

Figure 1. App selection process flowchart (completed March 2016).



Focus on Youth

We made efforts to focus our study on apps actually used by youth, and this was done by further narrowing down the selection from the initial 1200 apps identified. Apps were characterized as available to and targeted toward minors if they generally did not require the use of money and did not facilitate interaction with unknown people. Specific exclusion criteria included apps that (1) encourage the use of money outside in-app purchases (eg, shopping, travel, or real-estate apps), (2) facilitate interaction with unknown people (eg, dating or ride-service apps), (3) are focused on tracking pregnancies or newborn development, or (4) serve as licensing keys that unlock premium features of other apps (only in the Google Play Store). Shopping apps included apps related to specific stores or corporations (eg, Kohl's, Walmart, or Amazon), buy and sell apps (eg, letgo or eBay), and coupon or discount apps (eg, Groupon). Shopping apps did not include subscription streaming services such as HBO Now or Netflix. Dating and ride-service apps, including Tinder and Uber, were omitted because interaction with strangers is discouraged for youth.

Pregnancy and newborn development tracking apps were omitted because having and raising children is less common among teenagers and youth. A total of 96 apps were omitted. All other apps were included.

Reliability

To determine the reliability of the exclusion criteria, a second rater who had not seen the original list of 1200 apps applied the exclusion criteria to a random sample of 120 apps (30 per app type—Apple Free, Apple Paid, Google Play Free, and Google Play Paid). Out of the 120 apps, there was disagreement on only one app, yielding a kappa statistic of .94 ($P<.001$), which demonstrates high interrater agreement [17]. After discussion, the 2 raters came to consensus on the one app of disagreement and included it in the sample as "available to youth."

For the analysis of the apps, in each of the four app types, the highest ranked 30 apps, representing 10.00% (120/1200) of the apps, were reviewed for availability of a privacy policy. A total of 120 apps were considered a feasible number of privacy policies to analyze using a readability calculator. Of these 120 apps, 21 were available in both the Apple and Google stores and were analyzed only once. Out of the final 99 apps, 24 apps did not have privacy policies, and 11 apps had identical privacy policies because of those apps being products of the same developer. This left a total of 64 unique documents for our final readability analysis. Privacy policies of apps were found either via direct link to the privacy policy from the respective app store or from a link to the website of the app developer.

Readability Analysis

Comprehensibility was measured as "readability," or the ease of understanding the given text. Readability was used as a measure of comprehensibility, as it provides an unbiased numerical value reflective of comprehensibility. Readability statistics of privacy policies for apps from the Apple and Google Play app stores were calculated using a Web-based readability calculator and analyzed. The average RGL was then compared with the average RGL of adults in the United States. Notably, there are no standards or guidelines for the readability of mobile app privacy policies, so the readability statistics were also compared with the PPR TF. The PPR TF is a set of criteria that measure how technology affects patient privacy. These criteria were developed by the Coalition for Patient Privacy, in collaboration with others, to offer suggested standards on how patient privacy can be protected.

The 64 privacy policies were entered into a Web-based readability calculator, the Readability Test Tool (WebpageFX, Inc, Harrisburg, PA) [18], which is one of multiple free resources that calculate readability. Before selecting this tool, privacy policies were entered into multiple Web-based calculators. As most tools were found to produce fairly consistent results, the Readability Test Tool was used because of its simple user interface.

Statistics collected from the readability calculator were word count, Flesch reading ease, Flesch-Kincaid RGL, Gunning-Fog RGL, simplified measure of Gobbledygook (SMOG) RGL, sentence count, and number of complex words. Flesch reading ease computes a score on a scale from 0 to 100 with higher numbers representing greater reading ease. Flesch-Kincaid, Gunning-Fog RGL, and SMOG RGL are calculated by taking into account the sentence length and average word length. Gunning-Fog uses the average word length to determine the percentage of complex words or words with greater than three syllables. SMOG RGL typically overestimates the RGL of the text, and Flesh-Kincaid typically underestimates RGL. For a more accurate metric, RGL was calculated as the average of Flesch-Kincaid RGL, Gunning-Fog RGL, and SMOG RGL (Table 1).

Data Analysis

Mean RGL of the 64 apps was compared with the average adult reading level in the United States and to the PPR TF recommended RGL of 12.0. The Flesch reading ease score was compared with the PPR TF recommended reading ease score of 45.0. Apps were also divided into three broad app categories (entertainment, social networking, and utility) based on app store classifications. Entertainment apps included games, music, and video apps (eg, Angry Birds, Spotify, and Netflix). Social networking apps were categorized as such by the app stores and included messaging services associated with social networking (eg, Snapchat, Facebook Messenger, and Instagram). Utility apps encompassed all apps for general use and apps that did not fit into the other two categories (eg, flashlight, word processing, or email apps). RGL of the three categories were compared using a one-way analysis of variance (ANOVA). All reported *P* values are uncorrected.

Table 1. Flesch-Kincaid, Gunning-Fog, simplified measure of Gobbledygook (SMOG), and average reading grade levels (RGLs) for all apps included in the analysis. The average reading level column is the average of Flesch-Kincaid, Gunning-Fog, and SMOG RGLs.

App name	Average reading level	Flesch-Kincaid reading level	Gunning-Fog	SMOG[a]
Disney Build It: Frozen	17.1	16.8	19.4	15
Subway surfers	15.9	16.1	18.2	13.6
Nova Launcher Prime	15.6	15.8	16.9	14.1
Monument Valley	15.5	15.9	16.6	14
WhatsApp	15.5	16.2	17.4	13
Du Battery Saver and phone charger	15.2	14.5	17.9	13.3
Netflix	14.9	14.6	17.2	13
Grand Theft Auto: San Andreas	14.7	14.5	16.4	13.1
Mobile Strike	14.6	14.6	16.6	12.5
Pages	14.2	13.8	16.2	12.6
Terraria	14.2	13.7	16.3	12.5
Faily brakes	14.1	13.8	16.1	12.3
Pandora	14.1	13.9	16	12.5
Rolling Sky	14	13.4	16.4	12.3
Stick Texting: The Emoji Killer	13.9	13.8	15.7	12.4
Gmail	13.8	13.5	15.9	11.9
Assassin’s Creed Identity	13.7	13.8	14.8	12.5
Minecraft: Story Mode	13.7	13.5	15.6	12
Angry Birds	13.6	13.3	15.2	12.4
NBA 2K16	13.5	13.3	15	12.2
Candy Crush Jelly Saga	13.4	13	15	12.2
FaceSwap Live Lite	13.4	13.3	14.5	12.3
Ultimate Guitar Tabs and Chords	13.3	13	14.8	12
Twitter	13.2	13.2	15.2	11.2
Agar.io	12.9	12.3	15.3	11.3
Hitman: Sniper	12.9	12.7	14.2	11.7
Kimoji	12.9	12.6	14.7	11.5
Spotify Music	12.8	12.5	14.3	11.7
VivaVideo Pro	12.8	12.3	14.8	11.2
Facetune	12.7	12.7	13.9	11.6
Heads Up	12.7	12.3	14.5	11.3
Swype keyboard	12.7	12.5	13.9	11.6
Fishdom: Deep Dive	12.6	12.4	14.4	11.1
Game of Life Classic Edition	12.6	12.4	14.3	11.2
Geometry Dash	12.6	12.2	14.4	11.3
Power Clean: Optimize cleaner	12.6	12.4	13.9	11.5
Snapchat	12.5	12.2	14.6	10.8
Super Bright LED Flashlight	12.5	12.3	13.5	11.8
Clash Royale	12.4	11.8	14.4	11.1
Plague Inc	12.4	12.6	13.1	11.6
Sleep Cycle Alarm Clock	12.4	12.4	13.2	11.6

App name	Average reading level	Flesch-Kincaid reading level	Gunning-Fog	SMOG[a]
Bloon TD 5	12.3	12.3	14.1	10.6
Facebook	12.3	11.8	14.3	10.8
Instagram	12.3	12.1	13.9	11
Akinator the Genie	12.2	12.3	12.8	11.4
YouTube	12.2	11.7	14.4	10.6
Please Don't Touch Anything	11.9	11.6	12.8	11.2
Musical.ly	11.8	11.5	12.7	11.2
ZEDGE	11.8	11.4	13.4	10.7
Kik	11.7	11.3	13.3	10.6
PianoTiles 2	11.6	11.4	12.6	10.9
Dragon Land	11.4	10.7	12.8	10.6
Kika Emoji Keyboard	11.3	11	12.9	10
NeoMonsters	11.3	10.8	12.8	10.3
Pinterest	11.3	10.8	12.9	10.2
Toca Lab	11.2	11	12.6	10
Afterlight	10.8	10.4	11.9	10.2
Minecraft pocket edition	10.7	10.2	11.9	10.1
Badland 2	10.5	9.7	12.3	9.4
True Skate	10.2	10	11.4	9.1
Pocket Casts	10	9.2	11.5	9.3
SuperPhoto Full	10	9.4	11.7	9
The Room Three	10	9.3	11.7	9.1
Papa's Freezeria To Go	8.5	8.6	8.8	8.2

[a]SMOG: simplified measure of Gobbledygook.

Results

Readability

The privacy policies reviewed in our analysis had a mean length of 2425 words (standard deviation [SD] 1965) and ranged from 140 to 8290 words (Tables 2 and 3 and Figure 2). Privacy policies had a mean RGL of 12.78 (SD 1.611; Tables 2 and 3 and Figure 3). The correlation between privacy policy length and RGL was not statistically significant (r=.2452, P>.05, N=64). The mean Flesch reading ease was 42.73 (SD 6.991).

Policy Readability Versus Recommended Standards

Importantly, none of the discovered privacy policies had an RGL below the average adult RGL in the United States of 8.0 (Figure 3). Privacy policies also had an average Flesch reading ease of 42.73 (SD 6.991), which is lower (ie, less readable) than the 45.0 recommended reading ease by the PPR (P<.05; Figure 3). The average RGL of 12.78 is similar to the PPR TF recommended RGL of 12.0.

App Category Comparisons

The readability of policies from 30 free apps and 34 paid apps were compared. Free apps had an average RGL of 13.09 (SD 1.304), and paid apps had an average RGL of 12.51 (SD 1.815). Data are shown in Table 2 and illustrate no significant differences between free and paid apps on any of the metrics examined (P>.05). Apps were also divided into three broad categories (entertainment, social networking, and utility), as previously described. When privacy policies from these apps were compared as a function of category, we observed a significant difference in word count between the categories (Table 3), with social networking having the highest word count and utility the lowest. There were, however, no significant differences in average RGL.

Table 2. Mean readability statistics. Free versus paid: comparison of mean reading grade level (RGL), mean word count, and mean reading ease between free and paid apps from both the Android and Apple markets. The *P* values for the *t* tests between the two app types show that there is no significant difference between mean RGL, word count, and reading ease.

Statistic	All apps[a]	Free apps	Paid apps	*P* value
N	64	30	34	--
Mean RGL[b]	12.78	13.09	12.51	.15
Mean word count	2425	2355	2487	.79
Mean Flesch reading ease	42.73	42.3	43.1	.65

[a]Column summarizes results for all apps included in the analysis; it was not included in the significance test for the *P* value in the last column.

[b]RGL: reading grade level.

Table 3. Mean readability statistics. Entertainment versus social networking versus utility: comparison of mean reading grade level (RGL), mean word count, and mean reading ease between entertainment, social networking, and utility apps. The *P* values for the analysis of variance (ANOVA) tests show that there is no significant difference in the mean RGL and reading ease between the app categories, but there is a significant difference in mean word count.

Statistic	All apps[a]	Entertainment	Social networking	Utility	*P* value
N	64	44	7	13	--
Mean RGL[b]	12.78	12.84	12.7	12.62	.93
Mean word count	2425	2546	3493	1038	.02
Mean Flesch reading ease	42.73	42	46.46	43.37	.31

[a]Column summarizes results for all apps included in the analysis; it was not included in the significance test for the *P* value in the last column.

[b]RGL: reading grade level.

Figure 2. Privacy policy word count (N=64 apps). The average word count of the privacy policies was 2425 words. The Game of Life Classic Edition had the highest word count at 8290 words, and Plague Inc had the lowest word count at 140 words.

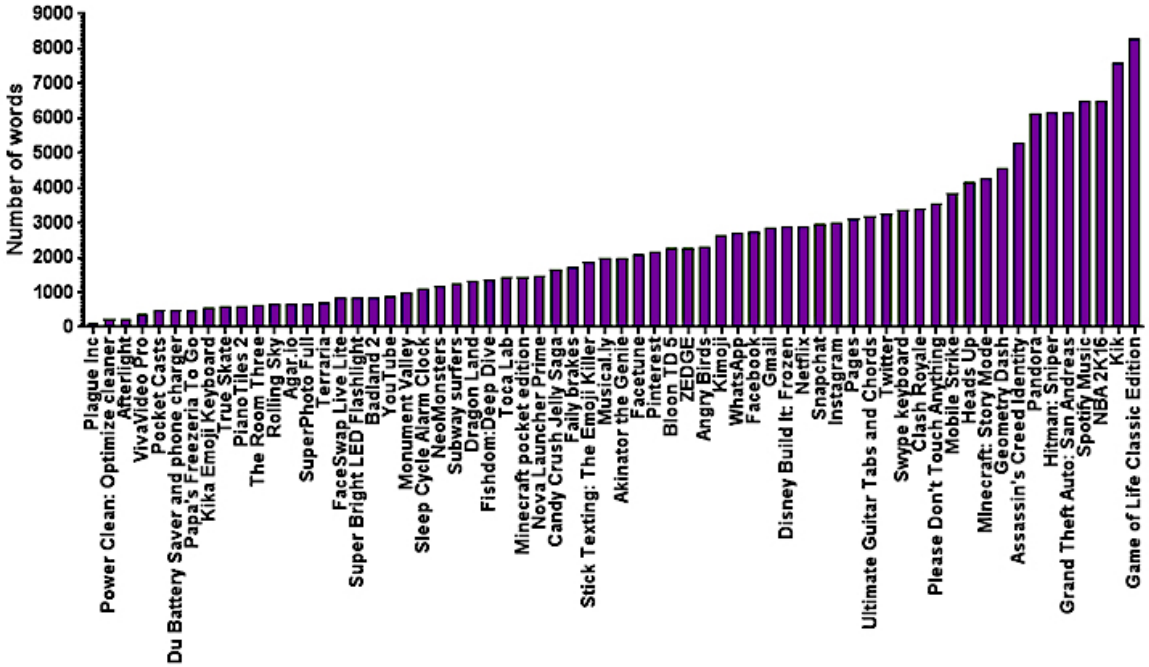
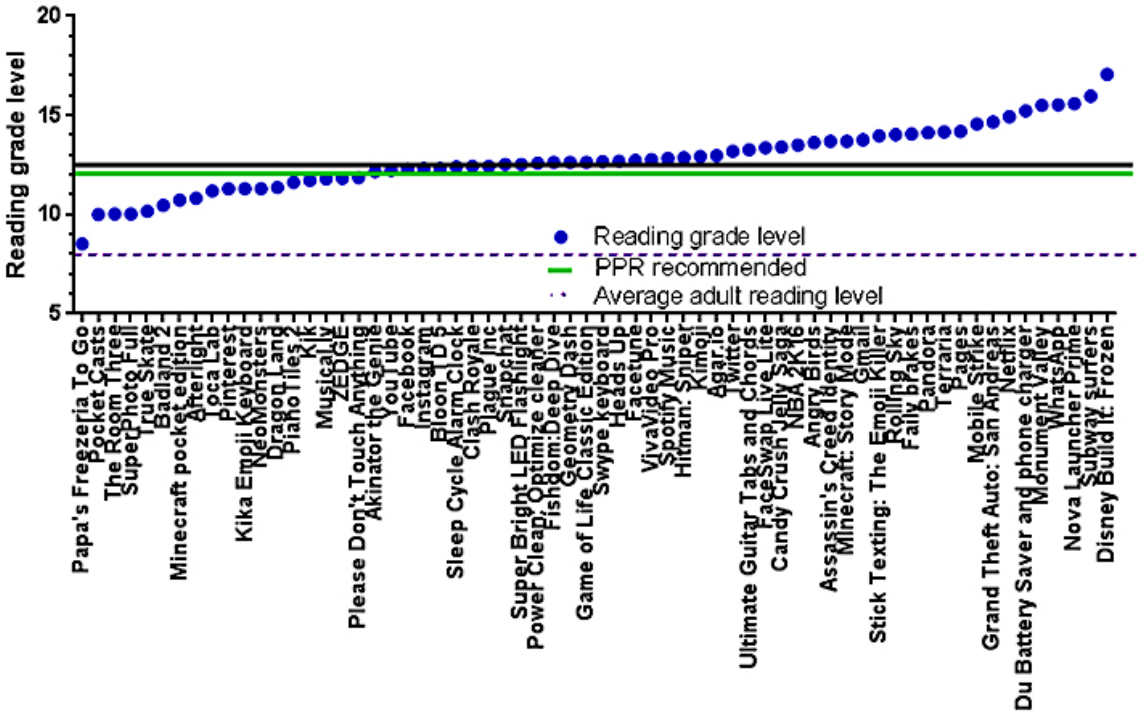


Figure 3. Privacy policy reading grade level (RGL; N=64). The RGL was an average of the Flesch-Kincaid, Gunning Fog, and simplified measure of Gobbledygook (SMOG) RGLs. The mean RGL of all the apps was 12.78, which is equivalent to a freshman in college. This average level is also higher than the Patient Privacy Rights (PPR) recommended RGL of 12.00 and higher than the US average adult RGL of 8.00. In terms of the individual apps, the highest RGL was for Disney Build It: Frozen at 17.07, which is equivalent to a graduate student reading level. The lowest RGL was for Papa's Freezeria To Go at 8.53.



Discussion

Principal Findings

Analysis of privacy policies for 64 popular apps targeted toward youth revealed an average reading level of 12.78 or the equivalent RGL of a first year college student. Although this RGL is similar to the reading level recommended by the PPR TF, it is well above the average reading level of adults in the United States. These findings are similar to those from a 2015 study (Sunyaev et al), which noted that app developers and companies are not transparent about their privacy practices through their privacy policies [4]. Although users must agree to app privacy policies to access digital tools and products, these agreements are not comprehensible by the average adult, let alone youth. Because companies often collect, use, and sell users' personal information, it is concerning that agreements describing and governing these activities are not accessible to most users. We propose that stakeholders, including pediatricians and other health care professionals, could play a role in educating youth and their guardians about the use of Web-based services and potential privacy risks, including the unintentional sharing of personal data. However, considering the complexities of privacy policy agreements, there may be a need for further tools and training to help such stakeholders, including health care workers, understand, navigate, and educate others about Web-based privacy and Internet safety.

Most parents are concerned about their child's safety on the Internet. Whereas many have taken steps to protect their child's safety while using the Web, such as through discussions with their children, it is often difficult for parents to know how their child's privacy is protected on the Internet [19]. About 40% of parents of Internet users have read the privacy policies of the apps that their children are using. Previous studies that have assessed privacy policies of mobile apps have concluded that college-level literacy is required to comprehend the text of privacy statements [20]. Likewise, our study reached similar conclusions even though the apps selected for analysis were specifically directed toward children and teenagers. Apps that are available to teenagers should have privacy statements that teenagers can understand, and apps that are available to children should have privacy statements that are accessible by their parents or guardians. To be COPPA compliant, apps and websites should post a policy regarding their privacy practices so that parents are aware of how information is collected and used, and these policies must be readable and comprehensible [21].

Results from a 2013 study conducted by the Pew Research Center show that 70% of teen Internet users do seek out advice about their Internet safety. Many teenagers turn to friends, peers, or their parents for advice about privacy settings on Web applications. The results of the Pew study also show that teenagers of all racial and socioeconomic backgrounds seek advice about Internet safety, but white teenagers are more likely than black or Hispanic teenagers to talk to their parents about Web privacy. Youth should have a trusted adult they can consult when considering privacy expectations with their Web presence. By having privacy policies written so that youth can understand them, children and teenagers are afforded a sense of autonomy over their Internet practices. They will be able to make informed

decisions about what kind of privacy settings they desire on their Web-based accounts, and they can discuss these privacy settings and their safety with a trusted adult [9].

Much of the inaccessible language in privacy policies stems from legal terminology used by corporations to protect themselves from potential liability. We identified excerpts from privacy policies in our study with the highest RGL (Table 4). Use of terms such as "cookies" and "third-party site" may contribute to comprehension difficulties, as well as complex phrases that use other jargon not in common parlance. It is well known that many users do not read privacy statements when they do download an app, and one possible reason for this may be the fact that they are difficult to comprehend. A potential solution is to require app developers to have versions of their privacy statements that translate the legal terminology in a way that is easy to understand. For example, Twitter's privacy policy includes one sentence "tips" that summarize different sections of the policy [22]. These tips are short and easy to read and allow users to better understand how their personal information is being used.

We noted that even the PPR TF criteria that was used as a base of comparison for readability in this study has recommended standards that are too difficult for the average adult in the United States to comprehend, as they recommend a RGL of 12.0. We recommend that a new set of guidelines for privacy policies target the average adult in the United States, with an average RGL of 8.0 or lower, a Flesch reading ease score of 70 or higher, and a word count of less than 500 words. These standards would also be understood by most high school students, allowing teenagers to read and comprehend privacy policies for the apps they download and potentially gain a better understanding of how their personal data are collected, used, and potentially sold to third parties.

The complexity and thus incomprehensibility of privacy policies poses a serious Internet safety concern for the youth in particular. A recent study on digital monitoring activity among teenagers shows that most parents do talk to their teenage children about appropriate Web behavior and what they should share on the Internet; however, most parents do not have these talks as frequently as they speak to their children about offline behavior [23]. With the increasing use of Web-based applications in entertainment, education, and social networking, young people are making more and more information available over the Web, potentially leading to harmful consequences.

Introducing educational curricula in schools about Web-based safety and increasing exposure to safe Internet practices may be an avenue to explore empirically. These curricula could provide children and adolescents with the tools they need to understand privacy risks and make choices about how their personal data are stored and shared over the Internet. Such resources are particularly important for older teenagers, who are less likely than younger children to involve their parents or ask for advice about Web privacy [9]. Indeed, teenagers are often already in the position of making their own choices about their behavior and practices in Web-based and digital environments. Web-based safety programs, such as the one developed by Common Sense Education, allow teachers to tailor their curricula to specific grade levels to make Internet safety relevant to minors of different ages [24].

Table 4. Sample text from privacy policies with highest reading grade level (top 5).

App name	Word count	Reading grade level	Sample text from privacy policy
Disney Build It: Frozen	2880	17.07	"We collect...Usage, viewing and technical data, including your device identifier or IP address, when you visit our sites...or open emails we send."
			"We acquire information from other trusted sources..."
Subway Surfers	1272	15.97	"We log information about your use of the App..."
			"...if you log into the App using a third-party site or platform such as Facebook, we may access information about you from that site or platform..."
			"We may allow third parties to serve contextual advertisements and provide analytics services in connection with the App. These entities may use various identifiers to collect information..."
Nova Launcher Prime	1487	15.60	"Information collected automatically from this Application (or third party services employed in this Application), which can include: the IP addresses or domain names of the computers utilized by the Users who use this Application...the country of origin..."
WhatsApp	2701	15.53	"WhatsApp will periodically access your address book or contact list on your mobile phone..."
			"WhatsApp uses both session cookies and persistent cookies. A persistent cookie remains after you close your browser..."
Monument Valley	984	15.50	"For operation and maintenance purposes, this Application and any third party services may collect files that record interaction with this Application (System Logs) or use for this purpose other Personal Data (such as IP Address)."
			"This Application does not support "Do Not Track" requests."

Given the ubiquitous nature of Web-based applications and the increasing frequency of use among children and adolescents, combined with the potential for harm if these are used inappropriately, health care providers may need to consider how to address these harms in the context of their overall care of underage patients. Using clinicians as a vehicle for counseling patients on privacy and app safety practices would be analogous to the ways in which health professionals play an important role in informing patients about practices to promote a healthy lifestyle (eg, physical activity and nutrition). For example, health care providers who interact with youth (eg, orthodontists, dentists, or pediatricians) can leverage their access to youth to share information about safety practices to enhance protection of youth in Web-based settings. However, to do that, a systematic approach to document the need for and, subsequently, appropriate guidelines directed to the clinician, would be needed.

Conclusions

Overall, Internet safety has increasingly become a public health issue. Whereas parents may have the primary responsibility for Internet safety education [8], the literature documents research findings that underscore the expertise required to understand privacy policies. The AAP has posted a guide on their website to assist parents in opening a dialogue to talk to their kids about Internet safety and social media [10]. Social networking features have become increasingly prevalent in apps—even apps that are not directly associated with social media are often linked to Facebook accounts or have the option to share on social networking. This expansive network increases opportunities for exposure to cyberbullying or material that is unsuitable for minors, which can lead to mental health and safety issues in the pediatric population [11]. Until there are clear standards for pediatricians and other health care providers specific to privacy and app safety education, they can assist by sharing information about available tools and educational resources.

Finally, institutional resources should be developed to help health professionals fulfill this role. An example of this is the AAP policy statement “Media Use in School-Aged Children and Adolescents” [25] that specifically highlights the privacy risks of social media and other Web-based activities and recognizes pediatricians’ role in helping parents set rules for Web-based activities and mentor their children about Web safety. Although the AAP tools are a good beginning, there is a need for further tools and training to help health care workers understand, navigate, and educate others about Web-based privacy and Internet safety. Overall, there is evidence that youth are concerned about maintaining their privacy, so training pediatricians and other health care providers to address privacy concerns with their patients will provide an additional safe place to ask questions and open a dialogue about Internet safety.

Acknowledgments

The authors thank Joshua Quiroz for his assistance with aspects of this project. The authors also thank Kathryn Montgomery, Mark Hochhauser, and Kevin Patrick for their comments on study design and interpretation. This work was supported by a grant from the National Human Genome Research Institute “Impact of Privacy Environments for Personal Health Data on Patients” (PI: CB, R01 HG008753, 2015-2018).

Conflicts of Interest

None declared.

References

1. Statista. Number of apps available in leading app stores as of July 2015 URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2017-03-01] [WebCite Cache ID 6oeJ6SFSh]
2. Lane N, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell A. A survey of mobile phone sensing. IEEE Commun Mag 2010 Sep 02;48(9):140-150. [doi: 10.1109/MCOM.2010.5560598]
3. FTC. 2013 Feb. Mobile privacy disclosures: building trust through transparency URL: https://www.ftc.gov/reports/mobile-privacy-disclosures-building-trust-through-transparency-federal-trade-commission [accessed 2017-09-18] [WebCite Cache ID 6tZcvUM92]
4. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
5. Democratic Media. 2016 Jun 26. The new children's online privacy rules: what parents need to know URL: https://www.democraticmedia.org/sites/default/files/CDDCOPPAParentguideJune2013.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZdWLYy9]
6. Harris K. OAG. 2014 May. Making your privacy practices public: recommendations on developing a meaningful privacy policy URL: https://oag.ca.gov/sites/all/files/agweb/pdfs/cybersecurity/making_your_privacy_practices_public.pdf [accessed 2017-09-18] [WebCite Cache ID 6tZeFKaJr]
7. Zang J, Dummit K, Graves J, Lisker P, Sweeney L. Who knows what about me? a survey of behind the scenes personal data sharing to third parties by mobile apps. Technology Science 2015 Oct 30:1-53 [FREE Full text]
8. Moreno MA, Egan KG, Bare K, Young HN, Cox ED. Internet safety education for youth: stakeholder perspectives. BMC Public Health 2013 Jun 5;13(1):543 [FREE Full text] [doi: 10.1186/1471-2458-13-543] [Medline: 23738647]

9. Lenhart A, Madden M, Cortesi S, Gasser U, Smith A. Pew Research Center. 2013 Aug 15. Where teens seek online privacy advice URL: http://www.pewinternet.org/2013/08/15/where-teens-seek-online-privacy-advice/ [accessed 2017-09-18] [WebCite Cache ID 6tZfUGixi]
10. American Academy of Pediatrics. 2013 May 31. Talking to kids and teens about social media and sexting URL: https://www.aap.org/en-us/about-the-aap/aap-press-room/news-features-and-safety-tips/Pages/Talking-to-kids-and-Teens-about-social-media-and-sexting.aspx [accessed 2017-07-27] [WebCite Cache ID 6sGERsdnF]
11. Diomidous M, Chardalias K, Magita A, Koutonias P, Panagiotopoulou P, Mantas J. Social and psychological effects of the internet use. Acta Inform Med 2016 Feb;24(1):66-68 [FREE Full text] [doi: 10.5455/aim.2016.24.66-68] [Medline: 27041814]
12. Livingstone S, Carr J, Byrne J. Cigionline. 2015 Nov 02. One in three: Internet governance and children's rights URL: https://www.cigionline.org/publications/one-three-internet-governance-and-childrens-rights [accessed 2017-09-18] [WebCite Cache ID 6tZfrMwgq]
13. Chia PH, Yamamoto Y, Asokan N. Is this app safe? a large scale study on application permissions and risk signals. New York, NY: International World Wide Web Conference Committee (IW3C2); 2012 Apr Presented at: Proceedings of the 21st International conference on World Wide Web; April 16-20, 2012; Lyon, France p. 1-10.
14. Peel D. Privacy Trust Framework. Patient Privacy Rights 2013 Feb 27:1-13 [FREE Full text] [doi: 10.2139/ssrn.2231667]
15. Apple Inc. URL: https://itunes.apple.com/us/genre/ios/id36?mt=8 [accessed 2017-03-01] [WebCite Cache ID 6oeK5W1R1]
16. Google Inc. URL: https://play.google.com/store/apps?hl=en [accessed 2017-09-18] [WebCite Cache ID 6tZh9INWW]
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
18. WebpageFX Inc. Readability test tool URL: http://www.webpagefx.com/tools/read-able/ [accessed 2017-03-02] [WebCite Cache ID 6oeLNVXdm]
19. Children's Commissioner. London; 2017 Jan. Growing up digital: a report of the growing up digital taskforce URL: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2017/06/Growing-Up-Digital-Taskforce-Report-January-2017_0.pdf [WebCite Cache ID 6tZhWccDQ]
20. McDonald AM, Reeder RW, Kelley PG, Cranor LF. A comparative study of online privacy policies and formats. In: Privacy Enhancing Technologies. Berlin, Heidelberg: Springer; 2009:37-55.
21. FTC. 2015. Complying with COPPA: frequently asked questions URL: https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions [accessed 2017-03-02] [WebCite Cache ID 6oeLGfMMv]
22. Twitter Inc. Twitter privacy policy URL: https://twitter.com/en/privacy [accessed 2017-09-18] [WebCite Cache ID 6tZiJXV98]
23. Anderson M. Pew Research Center. 2016 Jan 07. Parents, teens, and digital monitoring URL: http://www.pewinternet.org/2016/01/07/parents-teens-and-digital-monitoring/ [accessed 2017-09-18] [WebCite Cache ID 6tZieU27k]
24. Common Sense. Digital citizenship URL: https://www.commonsense.org/education/digital-citizenship [accessed 2017-03-02] [WebCite Cache ID 6oeLaSLQH]
25. Council on Communications and Media. Media use in school-aged children and adolescents. Pediatrics 2016 Nov;138(5):1-8 [FREE Full text] [doi: 10.1542/peds.2016-2592]

Abbreviations

AAP: American Academy of Pediatrics
ANOVA: analysis of variance
COPPA: Children's Online Privacy Protection Act
CalOPPA: California Online Privacy Protection Act
FTC: Federal Trade Commission
PPR TF: Patient Privacy Rights' Trust Framework
RGL: reading grade level
SD: standard deviation
SMOG: simplified measure of Gobbledygook

Edited by G Eysenbach; submitted 02.03.17; peer-reviewed by J Bender, J Li, P Boisrond, A Cyr; comments to author 05.06.17; revised version received 01.08.17; accepted 23.08.17; published 04.01.18.

Please cite as:
Das G, Cheung C, Nebeker C, Bietz M, Bloss C
Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability
JMIR Mhealth Uhealth 2018;6(1):e3
URL: http://mhealth.jmir.org/2018/1/e3/
doi:10.2196/mhealth.7626
PMID:29301737

Original Paper

Hospital-Owned Apps in Taiwan: Nationwide Survey

Hao-Yen Liu[1], MD; Wui-Chiang Lee[2,3], MHS, PhD, MD; Ying-Chou Sun[4], MD, PhD; Jun-Jeng Fen[5], MSc; Tzeng-Ji Chen[1,3], Dr Med; Li-Fang Chou[6], MSc, Dr Sc Pol; Shinn-Jang Hwang[1,3], MD

[1]Department of Family Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[2]Department of Medical Affairs and Planning, Taipei Veterans General Hospital, Taipei, Taiwan
[3]School of Medicine, National Yang-Ming University, Taipei, Taiwan
[4]Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan
[5]Department of Information Management, Taipei Veterans General Hospital, Taipei, Taiwan
[6]Department of Public Finance, National Chengchi University, Taipei, Taiwan

Corresponding Author:
Wui-Chiang Lee, MHS, PhD, MD
Department of Medical Affairs and Planning, Taipei Veterans General Hospital
No 201, Section 2, Shipai Road, Beitou District
Taipei, 11217
Taiwan
Phone: 886 2 28216999
Fax: 886 2 28202190
Email: wclee@vghtpe.gov.tw

Abstract

Background: Over the last decade, the use of mobile phone apps in the health care industry has grown rapidly. Owing to the high penetration rate of Internet use in Taiwan, hospitals are eager to provide their own apps to improve the accessibility of medical care for patients.

Objective: The aims of this study were to provide an overview of the currently available hospital-owned apps in Taiwan and to conduct a cross-hospital comparison of app features.

Methods: In May 2017, the availability of apps from all 414 hospitals in Taiwan was surveyed from the hospital home pages and the Google Play app store. The features of the downloaded apps were then examined in detail and, for each app, the release date of the last update, download frequency, and rating score were obtained from Google Play.

Results: Among all the 414 hospitals in Taiwan, 150 (36.2%) owned Android apps that had been made available for public use, including 95% (18/19) of the academic medical centers, 77% (63/82) of the regional hospitals, and 22.0% (69/313) of the local community hospitals. Among the 13 different functionalities made available by the various hospital-owned apps, the most common were the doctor search (100%, 150/150), real-time queue monitoring (100%, 150/150), and online appointment scheduling (94.7%, 142/150) functionalities. The majority of apps (57.3%, 86/150) had a rating greater than 4 out of 5, 49.3% (74/150) had been updated at some point in 2017, and 36.0% (54/150) had been downloaded 10,000 to 50,000 times.

Conclusions: More than one-third of the hospitals owned apps intended to increase patient access to health care. The most common app features might reflect the health care situation in Taiwan, where the overcrowded outpatient departments of hospitals operate in an open-access mode without any strict referral system. Further research should focus on the effectiveness and safety of these apps.

(JMIR Mhealth Uhealth 2018;6(1):e22) doi:10.2196/mhealth.8636

KEYWORDS

hospitals; telemedicine; mobile apps; Taiwan; mHealth

Introduction

Access to Mobile Health

Over the past few years, the proportion of the global population with access to mobile phone technology has surged, and it will continue to increase further, rising from an estimated 36.7% in 2016 to 74.7% in 2021 [1]. Mobile technology has penetrated every aspect of daily life, including health care. Mobile health (mHealth) apps are software tools that provide users with access to health-related information, facilitate health management, and provide connections between health care providers and users. At present, there are an estimated 259,000 mHealth apps available, from a total of 58,000 mHealth publishers, on the two main mobile device operating systems [2], and the global market for mHealth technology is expected to grow at an annual rate of 28.6% [3].

Thus far, mHealth apps have been shown to improve health outcomes by, for example, improving patients' medication adherence and adherence to lifestyle modifications [4-6], enhancing the control of risk factors for chronic diseases [7], changing health behaviors relating to smoking cessation [8], or providing access to health care professionals for the purposes of clinical decision making and patient monitoring [9]. For health providers, previous studies have shown that mHealth interventions are cost-effective and/or cost saving [10]. Some of the mHealth apps developed by hospitals are designed to enhance patient access, to improve communications between patients and providers, or to establish the hospital's brand online. One report in 2015 found that 66 of the 100 largest hospitals in the United States have their own mHealth apps, but that only 2% of the patients seen by those hospitals currently use those apps [11]. A US survey in 2016 claimed 52% of hospitals currently use three or more connected health technologies including mobile apps for patient education or engagement [12]. In China, Internet hospitals provide innovative health care directly through Internet technologies such as websites and mobile apps. In a study of the 43 established Internet hospitals, 18 (42%) of the hospitals provide access to outpatient health care via mobile app [13]. Related studies have focused on the contributions from the use of mHealth apps to specific disease outcomes [14-16]. However, to the best of our knowledge there have been no previous studies, much less any nationwide studies, that have analyzed the features of the mHealth apps provided by hospitals.

The mHealth System in Taiwan

Taiwan ranks fifth among the nations of the world in terms of mobile phone penetration, with approximately 70.4% of the population owning a mobile phone [17]. When it comes to the availability of a 3G or better data signal, Taiwan is among the top 10 nations worldwide, with an availability rate of 93.87% [18]. According to a recent report, Google's Android system dominates the market in Taiwan, being the operating system used on 68% of mobile devices [19]. Under Taiwan's National Health Insurance (NHI) system, citizens are granted free access to any specialist, even without a referral [20]. There are multiple ways to schedule a health care appointment in Taiwan, including by phone, online, or by showing up in person. Accordingly, it has become increasingly common for hospitals to ensure their patient volumes by releasing their own mHealth apps allowing for appointment scheduling, among other features.

The aim of this study was to provide an overview of hospital-owned apps in Taiwan. To do so, we recorded basic information about these hospital-owned apps and analyzed their various features to understand their value to health care consumers, which may in turn reflect the current health care situation in Taiwan. The findings of the study may thus provide guidance to health care providers as they seek to improve their mobile strategies and expand their service offerings.

Methods

Data Collection

A total of 416 hospitals in Taiwan received government accreditation from 2012 to 2015 [21]. Hospitals are accredited by the Taiwan Joint Commission on Hospital Accreditation, which is supervised by the Ministry of Health and Welfare, and classified into three levels based on health care quality, medical teaching ability, clinical capabilities, and bed capacity. In our study, the three levels of the surveyed hospitals were academic medical centers, regional hospitals, and local community hospitals.

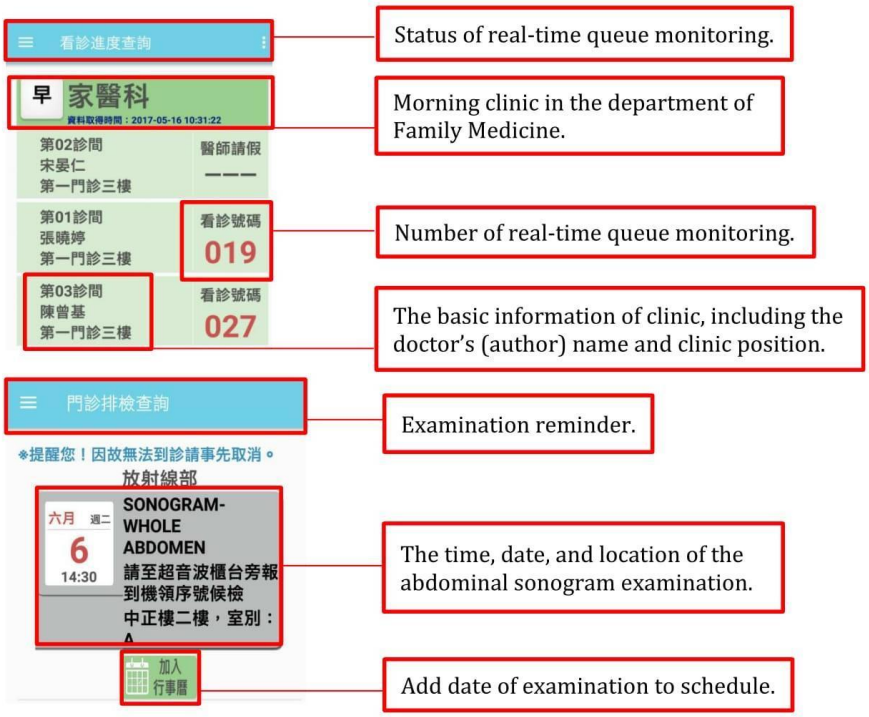
The locations of the hospitals were categorized according to the urbanization stratification of Taiwan's 368 townships developed by Taiwan's National Health Research Institutes [22]. Of the seven urbanization levels in that stratification, we defined levels 1 and 2 as urban, levels 3 and 4 as suburban, and levels 5 to 7 plus the isolated islands as rural. We excluded two hospitals located on remote islands because they are not included in the 368 townships.

Because of the higher prevalence of Android phone use in Taiwan and because information regarding the number of downloads was not available for iOS apps, we limited our study to Android system apps only. We used the name of each hospital to conduct a search for the given hospital's apps in the Google Play app store (Google Inc, Mountain View, CA, USA). We also created a user account and downloaded all the hospital-owned apps to a single Android phone (HTC One X9, HTC Corporation, Taoyuan, Taiwan) in May 2017. Apps were excluded if they were not available at that time or if the given app had no features relevant to health care availability for consumers. For the 414 hospitals investigated, our searches of the Google Play store identified a total of 150 apps.

Parameters of Consumer Interactions

We recorded basic data on the total number of reviews for all versions of each app using the data that was publically available via Google Play as of May 2017. Thus, the data analyzed was cross-sectional in nature. The average user ratings for the apps on a scale of one to five, categorized into three groups (4-5, 3.5-4, <3.5), were recorded. We also grouped the total number of downloads reported by Google Play into six different categories (100,000-500,000; 50,000-100,000; 10,000-50,000; 5000-10,000; 1000-5000; <1000), and recorded the given value for each app.

Figure 1. Screenshot of the real-time queue monitoring and examination reminder features of an app provided by Taipei Veterans General Hospital.



Apps Features Extraction and Review

A Microsoft Excel worksheet was constructed for the extraction of data regarding all the included apps. We listed features of these apps that are useful when patients are seeking health care and excluded features such as advertisements, games, and location. A total of 13 features of these apps were analyzed, including real-time queue monitoring (Figure 1), doctor search, appointment scheduling, appointment reminder, mobile payment, drug information, examination report, prescription refill reminder, personal health management, satisfaction and feedback, examination schedule, parking vacancy monitoring, and multilanguage features. Two investigators independently reviewed the features of each app, and the features were then cross-verified.

Statistical Analysis

A boxplot was constructed to present the rating scores for all the apps, which were then stratified by the three hospital types. For each boxplot of each hospital type, the bottom of the box represented the 25th percentile, the top of the box represented the 75th percentile, and the midline in the box represented the 50th percentile of the mean rating score of the app. The size of each box can provide an estimate of the rating score distribution of these apps.

Results

Characterization of the Hospitals

Out of the total analyzed sample of 414 hospitals, we identified 150 hospitals for which hospital-owned apps were available online, including 18 academic medical centers, 63 regional hospitals, and 69 local community hospitals. According to a 2016 report, daily outpatient visits in academic medical centers accounted for 28.81% (106,458/369,552) of the total hospital outpatient visits, followed by 30.79% (113,785/369,552) in regional hospitals and 34.07% (125,913/369,552) in local community hospitals [23]. All these hospitals provided Internet-based appointment scheduling systems through their home pages. Over the course of this study, all the Android apps developed by these hospitals appeared in the search results for the names of the different hospitals. All the academic medical centers (n=18) were located in urban regions. Of the regional hospitals (n=63), 68% (43/63) were located in urban regions, 30% (19/63) were located in suburban regions, and 2% (1/63) were located in rural regions. Of the local community hospitals (n=69), 48% (33/69) were situated in urban areas, 41% (28/69) were situated in suburban areas, and 12% (8/69) were situated in rural areas.

Distribution of Hospital-Owned Apps

As shown in Table 1, 150 of 414 hospitals had designed and made available their own apps, with those 150 including 95% (18/19) of the hospitals in the academic medical center group, 77% (63/82) of the hospitals in the regional hospital group, and 22.0% (69/313) of the hospitals in the local community hospital group. These results indicated notable differences among the three different types of hospitals. However, the proportions of hospitals with their own apps did not differ in terms of the three different urbanization levels.

Table 1. Proportion of apps (n=150) developed in all hospitals (N=414), stratified by accreditation type and urbanization level[a].

Hospital characteristics	Academic medical center (n=19), n/N (%)	Regional hospital (n=82), n/N (%)	Local community hospital (n=313), n/N (%)	Total (N=414), n/N (%)
Urban (n=253)	18/19 (95)	43/56 (77)	33/178 (18.5)	94/253 (37.2)
Suburban (n=135)	N/A[b]	19/25 (76)	28/110 (25.5)	48/135 (35.6)
Rural (n=26)	N/A[b]	1/1 (100)	8/25 (32)	9/26 (35)
Total (N=414)	18/19 (95)	63/82 (77)	69/313 (22.0)	150/414 (36.2)

[a]Values are hospitals with apps/all hospitals (percentages) unless otherwise indicated.

[b]N/A: not applicable.

Table 2. Content and features of hospital-owned apps (N=150).

Apps characteristics	Academic medical center (n=18), n (%)	Regional hospital (n=63), n (%)	Local community hospital (n=69), n (%)	Total (N=150), n (%)
Last update time				
2017	10 (56)	28 (44)	36 (52)	74 (49.3)
2016	6 (33)	20 (32)	22 (32)	48 (32.0)
2015 and before	2 (11)	15 (24)	11 (16)	28 (18.7)
Features				
Real-time queue monitoring	18 (100)	63 (100)	69 (100)	150 (100)
Doctor search	18 (100)	63 (100)	69 (100)	150 (100)
Appointment scheduling	18 (100)	59 (94)	65 (94)	142 (94.7)
Appointment reminder	5 (28)	20 (32)	17 (25)	42 (28.0)
Mobile payment	11 (61)	11 (17)	12 (17)	34 (22.7)
Drug information	8 (44)	14 (22)	12 (17)	34 (22.7)
Examination report	7 (39)	11 (17)	14 (20)	32 (21.3)
Prescription refill reminder	6 (33)	9 (14)	10 (14)	25 (16.7)
Personal health management	3 (17)	9 (14)	10 (14)	22 (14.7)
Satisfaction and feedback	3 (17)	3 (5)	9 (13)	15 (10.0)
Examination schedule	1 (6)	1 (2)	1 (1)	3 (2.0)
Parking vacancy monitoring	2 (11)	N/A[a]	1 (1)	3 (2.0)
Multilanguage	1 (6)	1 (2)	1 (1)	3 (2.0)

[a]N/A: not applicable.

Features of Apps

Of the 150 apps, all had real-time queue monitoring and doctor search capabilities. In terms of availability, those capabilities were followed by appointment scheduling (94.7%, 142/150), appointment reminder (28.0%, 42/150), drug information (22.7%, 34/150), mobile payment (22.7%, 34/150), prescription refill reminder (16.7%, 25/150), personal health management (14.7%, 22/150), and satisfaction and feedback (10.0%, 15/150) functions. Only three of the 150 apps offered content in multiple languages, including English, as well as parking vacancy monitoring and examination reminder functions. Table 2 summarizes the distribution of the 13 app features across the different hospital levels.

For the number of app downloads, the apps for 36.0% (54/150) of the hospitals had been downloaded 5000 to 10,000 times (Figure 2). Among the academic medical centers, half (9/18) of the corresponding apps had been downloaded 100,000 to 500,000 or 50,000 to 100,000 times. In contrast, 46% (32/69) of the local community hospital apps were downloaded fewer than 5000 times. Regarding the date of the most recent update, 49.3% (74/150) of the hospitals with apps had updated their respective apps in 2017. There were no significant differences in update date among the different hospital accreditation levels.

Figure 2. Download distribution of hospital-owned apps (N=150).

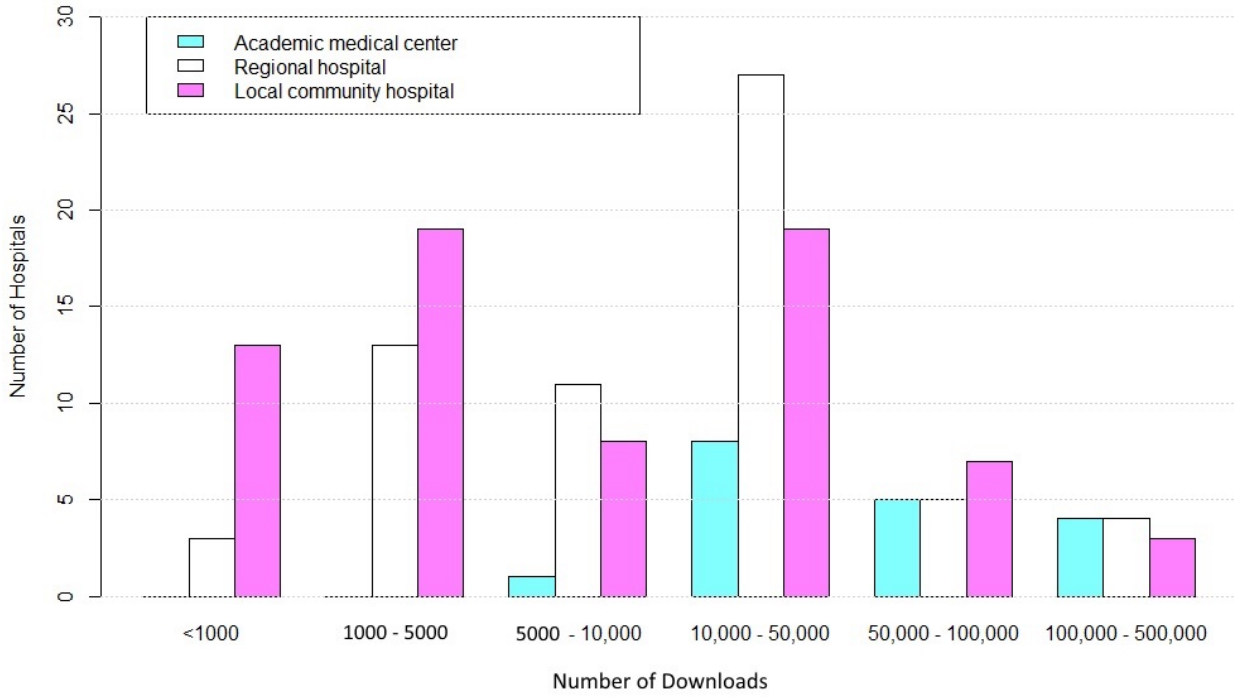
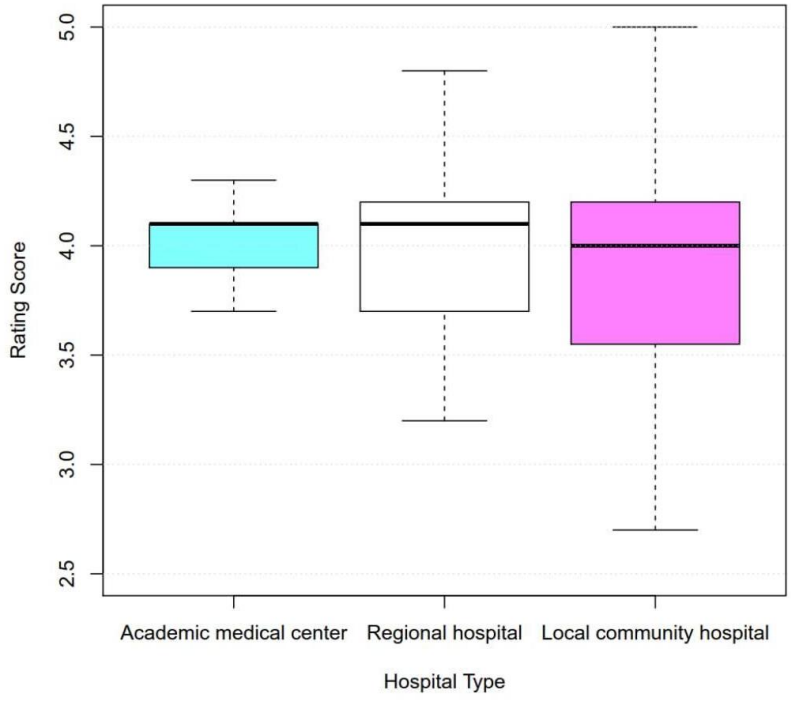


Figure 3. Users' rating scores by hospital type.



The majority of the apps (57.3%, 86/150) had received a high mean rating of between 4 and 5, 27.3% (41/150) had received a mean rating of between 3.5 and 4, and 14.0% (21/150) had received a poor mean rating (ie, mean rating <3.5). We also used the rating score data to illustrate the rating score distributions for the various apps according to the three different types of hospitals (Figure 3). The mean rating scores were 4.01 (SD 0.21) for the academic medical center apps, mean 3.99 (SD 0.42) for the regional hospital apps, and mean 3.89 (SD 0.70) for the local community hospital apps. The box in Figure 3 was largest for the local community hospital group apps, indicating that the ratings for those apps exhibited the largest degree of variation, followed by the box for the regional hospital group apps and the box for the academic medical center group apps.

Discussion

Principal Findings

The results of this study provide an overview of the distribution of hospital-owned apps in Taiwan as of May 2017. The differences in said distribution among the three types of hospitals and among the hospitals in the different types of regions might reflect the demands of the patients and the mobile penetration rates for the hospitals of the different types and regions. The key features of the identified mHealth apps were appointment scheduling, real-time queue monitoring, and doctor search functions, followed by appointment reminder, drug information, mobile payment, personal health management, satisfaction and feedback, and other functions. Approximately half (49.3%, 74/150) of the apps were updated in 2017, 36.0% (54/150) had been downloaded 10,000 to 50,000 times, and 57.3% (86/150) had a rating greater than 4 out of 5. All the available apps could be downloaded for free.

Distribution of Apps

This study found that the proportion of hospital-owned apps was significantly higher among hospitals of larger scale. The citizens of Taiwan can visit any level of hospital directly without a referral [20]. According to the 2015 annual report from the Ministry of Health and Welfare, the proportion of patient encounters in all hospitals was 36% (n=402,621 per day; 10.4%, 13.3%, and 9.3% in academic medical centers, regional hospitals, and local community hospitals, respectively) in 2015 [24,25]. Owing to greater financial, research, and educational resources, academic medical centers were often the preferred alternative for citizens in Taiwan, even though they typically charge higher co-payments and are more likely to be overcrowded. To remain competitive and ensure the maintenance of their incomes, these hospitals have aggressively increased access to care, leading to increased market share. In this survey, almost all the academic medical centers had made their own apps available, with the exception of one with an Internet-based appointment service only. Among the local community hospitals, fewer had their own apps and of those that did have apps, half of them had a number of downloads less than 5000. A possible reason for this might be the relatively low service volumes of local community hospitals compared to academic medical centers and regional hospitals. It was notable that we found similar proportions of hospitals with their own apps among regions with different urbanization levels. As such, whether there is a significant digital divide between hospitals in urban and rural areas requires further research.

Features of Apps

Appointment Reminder

Overall, nearly all the apps had the core features of real-time queue monitoring, appointment scheduling, and doctor search. In Taiwan, the average outpatient department visit rate has been reported to be up to 14 times per year per person, higher than the rates in other countries [20]. In the overcrowded outpatient clinics, which generally do not maintain waiting lists, these high-yield features of apps are particularly desired by patients seeking better access to appointments and the ability to monitor clinic queues regardless of time and place. More than one in four apps in the study also had an appointment reminder feature. Previous studies have reported that appointment reminder systems, such as short message service text messaging, can increase attendance at appointments and improve cancelation and rescheduling of unwanted appointments [26,27]. In addition, mHealth reminder systems have been reported to improve no-show rates and be more cost-effective than conventional strategies [28,29].

Medication Adherence

Poor medication adherence is particularly problematic among patients with chronic illnesses, and increasing adherence might have greater impact on health outcomes than specific medical therapies [30,31]. Owing to the large number of patients that they see, physicians in Taiwan usually spend less time on health counseling, including education regarding medication usage, than doctors in other countries [32]. In this study, we noticed that the prescription refill reminder and drug information features of apps were mainly aimed at improving patients' medication adherence and self-management. However, there is still a lack of evidence that mobile apps have a beneficial effect on medication adherence [33,34].

Mobile Payment

Mobile payment technology has gained widespread adoption in the restaurant and retail industries, but most health care providers still receive paper-based payments from patients. However, a report in 2016 found that mobile payments had doubled to 20% of all online payments in the preceding 3 years, and that consumer demand for convenience through online payment channels was increasing [35]. Accordingly, hospitals have begun to integrate mobile payment systems into their mHealth apps. In this study, approximately one-fifth of the identified apps provided users options for paying their medical bills via the digital platform. As such, instead of waiting in front of a clerk, patients can reduce unnecessary wait times in hospitals. For health care providers, such mobile payment options might give staff members more flexibility and availability by reducing the associated paper work.

Personal Health Record and Safety Concern

In this study, approximately one-fifth of the identified apps provided users with access to their examination reports and data. It is possible that patients might effectively improve their health outcomes and self-management by monitoring their personal health records via mHealth apps. On the other hand, patients might also struggle to interpret their medical data via the complex interfaces of apps [36]. Furthermore, it should also be noted that there are corresponding legal concerns and risks related to data protection [37]. For example, the personal record could be easily captured via entering the patient's ID number without further certification from interface of some apps. We also found no sufficient information or privacy policies in the app store descriptions regarding the guidelines for restricted content or the security of the data provided via the apps. Previous studies showed that mHealth app use might make widespread use of unsecured wireless network [38]. These concerns might lead to resistance to the use of mHealth apps

from stakeholders. According to one study about online health data, approximately 41% of US consumers have privacy and security concerns related to usage of mobile phone health devices [39]. In 2015, the US Food and Drug Administration released new guidance regarding the regulations for specific types of mHealth apps [40]. Increased use of mHealth apps could leave health data unsecured unless app developers keep improving the way they communicate and store data [38].

In the future, the developers of these mHealth apps should take regulation and risk assessment into consideration to ensure the safety of these apps.

Limitations

One possible limitation of this study is that this was a cross-sectional study. All the results regarding the parameters of apps may be out of date fairly soon because existing apps are updated and additional apps are continually released. Second, the search for apps was limited to the Android app store because the Android operating system has the highest market share; therefore, the results of this study might not be representative of the apps for the iOS and other platforms. Third, we could not precisely assess the rating scores of every app because some of the apps had only received a few reviews. In addition, assessing the quality of apps in terms of the star ratings given by users may not be the most reliable method of judging said quality [41]. Future studies are thus needed to provide a multidimensional measure for the quality assessment of mHealth apps. Fourth, in our research we did not investigate how these hospital-owned apps may increase accessibility, care quality, or patient satisfaction. Further discussions will be needed especially from administrators' points of view.

Conclusions

More than one-third of hospitals in Taiwan had their own mHealth app aimed at increasing patient access to health care. The most common app features might reflect the current health care situation in Taiwan, where the overcrowded outpatient departments of hospitals in the NHI system operate in an open-access mode without any strict referral system. Continued research is necessary to evaluate the beneficial effects on health outcomes contributed by these mHealth apps. Developers should design features in apps that can adequately address the demands of users and should focus seriously on issues relating to the regulation and safety of these mHealth apps.

Acknowledgments

The authors would like to thank Taipei Veterans General Hospital for administrative assistance and providing technical support to this project.

Authors' Contributions

HYL, TJC and WCL conceived and designed the investigations. YCS, JJF, LFC and SJH helped to perform the investigations and analyze the data. HYL wrote the manuscript. TJC and WCL revised the manuscript. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Proportion of apps (n=150) developed in all hospitals (N=414), stratified by accreditation type and urbanization level.a.

[JPG File, 88KB - mhealth_v6i1e22_app1.jpg]

Multimedia Appendix 2

Content and features of hospital-owned apps (N=150).

[JPG File, 158KB - mhealth_v6i1e22_app2.jpg]

References

1. Cisco. 2016 Mar 28. Cisco Visual Networking Index: global mobile data traffic forecast update, 2016–2021: white paper. URL: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html [accessed 2017-12-14] [WebCite Cache ID 6viPZQ9SG]
2. Research2guidance. mHealth App Developer Economics: The current status and trends of the mHealth app market. 2016. URL: https://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf [WebCite Cache ID 6sBAVPMR4]
3. BCC Research. 2017. Mobile health (mHealth) technologies and global markets. URL: https://www.bccresearch.com/market-research/healthcare/mobile-health-technologies-report-hlc162b.html [accessed 2017-07-23] [WebCite Cache ID 6sBAofX5n]

4. Fanning J, Mullen SP, McAuley E. Increasing physical activity with mobile devices: a meta-analysis. J Med Internet Res 2012;14(6):e161 [FREE Full text] [doi: 10.2196/jmir.2171] [Medline: 23171838]
5. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
6. Ernsting C, Dombrowski SU, Oedekoven M, Kanzler M, Kuhlmey A, Gellert P. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
7. McLean G, Band R, Saunderson K, Hanlon P, Murray E, Little P, et al. Digital interventions to promote self-management in adults with hypertension systematic review and meta-analysis. J Hypertens 2016 Apr;34(4):600-612 [FREE Full text] [doi: 10.1097/HJH.0000000000000859] [Medline: 26845284]
8. Ubhi HK, Kotz D, Michie S, van Schayck OC, Sheard D, Selladurai A, et al. Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. Addict Behav 2016 Jul;58:175-181 [FREE Full text] [doi: 10.1016/j.addbeh.2016.02.026] [Medline: 26950256]
9. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
10. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
11. Accenture. Losing Patience: Why healthcare providers need to up their mobile game. 2016. URL: https://www.accenture.com/t20151112T042615__w__/us-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_24/Accenture-Losing-Patience.pdf [WebCite Cache ID 6sBBpTtA4]
12. Healthcare Information and Management Systems Society. 2016 Mar 02. The 2016 HIMSS Connected Health Survey. URL: http://www.himss.org/2016-connected-health-survey [accessed 2017-10-12] [WebCite Cache ID 6uAPt5Kpp]
13. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet hospitals in China: cross-sectional survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]
14. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M, Herreros-González J. Mobile apps in cardiology: review. JMIR Mhealth Uhealth 2013;1(2):e15 [FREE Full text] [doi: 10.2196/mhealth.2737] [Medline: 25098320]
15. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
16. Nightingale R, Hall A, Gelder C, Friedl S, Brennan E, Swallow V. Desirable components for a customized, home-based, digital care-management app for children and young people with long-term, chronic conditions: a qualitative exploration. J Med Internet Res 2017 Jul 04;19(7):e235 [FREE Full text] [doi: 10.2196/jmir.7760] [Medline: 28676470]
17. Newzoo. Top countries by smartphone users and penetration. 2017. URL: https://newzoo.com/insights/rankings/top-50-countries-by-smartphone-penetration-and-users/ [accessed 2017-07-23] [WebCite Cache ID 6sBC15ULn]
18. OpenSignal. 2016. Global state of mobile networks (August). URL: https://opensignal.com/reports/2016/08/global-state-of-the-mobile-network [accessed 2017-07-23] [WebCite Cache ID 6sBCR14tT]
19. StatCounter. 2017. Mobile operating system market share in Taiwan. URL: http://gs.statcounter.com/os-market-share/mobile/taiwan [accessed 2017-07-23] [WebCite Cache ID 6sBDRfDEB]
20. Wu T, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. London J Prim Care (Abingdon) 2010 Dec;3(2):115-119 [FREE Full text] [Medline: 25949636]
21. Ministry of Health and Wealth, Republic of China. Hospitals qualified through accreditation during 2012 to 2015 in Taiwan. 2016. URL: http://www.mohw.gov.tw/dl-5592-cfbc7c75-4dde-4dd7-b6c9-4da90fedbf33.html [WebCite Cache ID 6viQMBzTb]
22. Liu CY, Hung YT, Chuang YL, Chen YJ, Weng WS, Liu JS, et al. Incorporating development stratification of Taiwan townships into sampling design of large scale health interview survey. J Health Man 2006 Jun 01;4(1):1-22 [FREE Full text]
23. Ministry of Health and Welfare, Republic of China. Average daily outpatient visits stratified by accreditation type in 2016. 2017 Jun 16. URL: http://dep.mohw.gov.tw/DOS/cp-1885-32574-113.html [WebCite Cache ID 6viSGaVi0]
24. Ministry of Health and Welfare, Republic of China. Annual reports and statistics of healthcare volume in 2015. 2016. URL: http://www.mohw.gov.tw/dl-19381-a5d2439b-3c44-4ef7-8c66-6c05ac1589bc.html [WebCite Cache ID 6viSst2aj]
25. Ministry of Health and Welfare, Republic of China. Trend of patient panel size in hospitals stratified by accreditation type From 2006 to 2020. 2006. URL: http://www.mohw.gov.tw/dl-2867-7b8536c5-1a8f-4374-b2c5-ad5dabbcb707.html [WebCite Cache ID 6viTKKa5a]
26. McLean S, Booth A, Gee M, Salway S, Cobb M, Bhanbhro S, et al. Appointment reminder systems are effective but not optimal: results of a systematic review and evidence synthesis employing realist principles. Patient Prefer Adher 2016 Apr:479. [doi: 10.2147/ppa.s93046]

27. Hallsworth M, Berry D, Sanders M, Sallis A, King D, Vlaev I, et al. Correction: stating appointment costs in SMS reminders reduces missed hospital appointments: findings from two randomised controlled trials. PLoS One 2015 Oct;10(10):e0141461 [FREE Full text] [doi: 10.1371/journal.pone.0141461] [Medline: 26484861]
28. Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 2016 Jan 14;16:13 [FREE Full text] [doi: 10.1186/s12913-015-1243-z] [Medline: 26769153]
29. Sims H, Sanghara H, Hayes D, Wandiembe S, Finch M, Jakobsen H, et al. Text message reminders of appointments: a pilot intervention at four community mental health clinics in London. Psychiatr Serv 2012 Feb 1;63(2):161-168. [doi: 10.1176/appi.ps.201100211] [Medline: 22302334]
30. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]
31. Lu W, Wen Y, Chen L, Hsiao F. Effect of polypharmacy, potentially inappropriate medications and anticholinergic burden on clinical outcomes: a retrospective cohort study. Can Med Assoc J 2015 Feb 02;187(4):E130-E137. [doi: 10.1503/cmaj.141219] [Medline: 25646290]
32. Lynn A, Shih T, Hung C, Lin M, Hwang S, Chen T. Characteristics of ambulatory care visits to family medicine specialists in Taiwan: a nationwide analysis. PeerJ 2015;3:e1145 [FREE Full text] [doi: 10.7717/peerj.1145] [Medline: 26290798]
33. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
34. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]
35. InstaMed. 2017. Trends in healthcare payments: seventh annual report 2016. URL: https://www.instamed.com/about/white-papers [accessed 2017-09-01] [WebCite Cache ID 6sBEOqwdv]
36. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2016 Oct 3:1 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
37. Lewis TL, Wyatt JC. mHealth and mobile medical apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210. [doi: 10.2196/jmir.3133] [Medline: 25223398]
38. He D, Naveed M, Gunter CA, Nahrstedt K. Security concerns in android mHealth apps. AMIA Annu Symp Proc 2014;2014:645-654 [FREE Full text] [Medline: 25954370]
39. Parks Associates. 2015 Aug 04. One-quarter of consumers have privacy concerns about using connected health devices. URL: http://www.parksassociates.com/blog/article/pr-aug2015-health-privacy [accessed 2017-12-14] [WebCite Cache ID 6viZFUUDJ]
40. Food and Drug Administration. Mobile medical applications: Guidance for Food and Drug Administration staff. 2015 Feb 09. URL: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf [WebCite Cache ID 6sBEjtyaX]
41. Yasini M, Marchand G. Mobile health applications, in the absence of an authentic regulation, does the usability score correlate with a better medical reliability? Stud Health Technol Inform 2015;216:127-131. [Medline: 26262024]

Abbreviations

NHI: National Health Insurance

Edited by G Eysenbach; submitted 03.08.17; peer-reviewed by CY Li, I Frederix, A Hall; comments to author 14.09.17; revised version received 03.11.17; accepted 16.11.17; published 16.01.18.

Please cite as:
Liu HY, Lee WC, Sun YC, Fen JJ, Chen TJ, Chou LF, Hwang SJ
Hospital-Owned Apps in Taiwan: Nationwide Survey
JMIR Mhealth Uhealth 2018;6(1):e22
URL: http://mhealth.jmir.org/2018/1/e22/
doi:10.2196/mhealth.8636
PMID:29339347

Original Paper

Improving Refill Adherence in Medicare Patients With Tailored and Interactive Mobile Text Messaging: Pilot Study

Rena Brar Prayaga[1], MA, JD; Erwin W Jeong[2], PharmD; Erin Feger[1], BA; Harmony K Noble[3], BA; Magdalen Kmiec[1], BA; Ram S Prayaga[1], MS

[1]mPulse Mobile, Inc, Encino, CA, United States

[2]Medicare Medication Therapy Management, Kaiser Permanente Southern California, Downey, CA, United States

[3]Regional Pharmacy Clinical Operations, Kaiser Permanente Southern California, Downey, CA, United States

Corresponding Author:
Rena Brar Prayaga, MA, JD
mPulse Mobile, Inc
16530 Ventura Blvd. Suite 500
Encino, CA,
United States
Phone: 1 888 678 5735
Email: rena@mpulsemobile.com

Abstract

Background: Nonadherence is a major concern in the management of chronic conditions such as hypertension, cardiovascular disease, and diabetes where patients may discontinue or interrupt their medication for a variety of reasons. Text message reminders have been used to improve adherence. However, few programs or studies have explored the benefits of text messaging with older populations and at scale. In this paper, we present a program design using tailored and interactive text messaging to improve refill rates of partially adherent or nonadherent Medicare members of a large integrated health plan.

Objective: The aim of this 3-month program was to gain an understanding of whether tailored interactive text message dialogues could be used to improve medication refills in Medicare patients with one or more chronic diseases.

Methods: We used the mPulse Mobile interactive text messaging solution with partially adherent and nonadherent Medicare patients (ie, over age 65 years or younger with disabilities) of Kaiser Permanente Southern California (KP), a large integrated health plan, and compared refill rates of the text messaging group (n=12,272) to a group of partially adherent or nonadherent Medicare patients at KP who did not receive text messages (nontext messaging group, n=76,068). Both groups were exposed to other forms of refill and adherence outreach including phone calls, secure emails, and robo-calls from December 2016 to February 2017.

Results: The text messaging group and nontext messaging group were compared using an independent samples t test to test difference in group average of refill rates. There was a significant difference in medication refill rates between the 2 groups, with a 14.07 percentage points higher refill rate in the text messaging group (P<.001).

Conclusions: The results showed a strong benefit of using this text messaging solution to improve medication refill rates among Medicare patients. These findings also support using interactive text messaging as a cost-effective, convenient, and user-friendly solution for patient engagement. Program outcomes and insights can be used to enhance the design of future text-based solutions to improve health outcomes and promote adherence and long-term behavior change.

(JMIR Mhealth Uhealth 2018;6(1):e30) doi:10.2196/mhealth.8930

KEYWORDS

patient activation; patient engagement; medication adherence; refill management; text messaging; interactive; NLP; Medicare; disease management; technology acceptability model

Introduction

Overview

Patient nonadherence affects 50% to 60% of chronically ill patients, and the cost of medication-related hospitalizations is $100 billion annually [1-3]. It is also associated with poor outcomes and progression of disease causing approximately 125,000 deaths and at least 10% of hospitalizations every year [4]. Seniors take an average of 7 medications per day, representing the highest number of prescribed medications for any age group [5].

Nonadherence is a major concern in the management of chronic conditions such as hypertension, cardiovascular disease, and diabetes where patients may discontinue or interrupt their medication for a variety of reasons. Patients are considered adherent when they take their medications (dose, time, frequency) as prescribed by their health care provider and as agreed to by the patient. Medicare populations adherence rates are often measured by pharmacy refill rates. The Centers for Medicare and Medicaid Services (CMS) uses the proportion of days covered (PDC), developed by Pharmacy Quality Alliance, to calculate adherence. Based on this, a patient who has a PDC rate of at least 80% is considered to be adherent.

Adherence is a particularly difficult problem among Medicare populations, and adherence rate is a key metric used by CMS to measure quality of a managed care plan. Approximately 32% of Medicare Part D patients are nonadherent to their diabetes, hypertension, and cholesterol medications [6]. Reasons for nonadherence may include side effects of the drug, cost of the drug, lack of perceived benefit, and/or forgetfulness.

Use of Mobile Technology for Adherence

Studies and surveys are finding that digital health is not reaching most seniors, especially where there are socioeconomic disparities [7]. Among seniors who are identified as tech-savvy, 70% of those surveyed believe it's important to be able to request prescription refills electronically, but fewer than half (46%) say they can do so today [8]. On researching mobile phone device ownership among seniors, we learned that while 78% of Americans aged 65 years and older own a mobile phone, only 34% own a smartphone [9,10]. We estimated smartphone ownership to be even lower among Medicare populations aged 65 years and older.

Text messaging using SMS (short message service) is ubiquitous, highly accessible, affordable, and commonly used across all income levels. It is also an effective channel for continuing to engage individuals in their health care once they leave the doctor's office. Interactive text dialogues provide an opportunity for patients and health plan members to tap into health care resources and get support for healthy behaviors and long-term behavior change. Several studies have found that text messaging may serve as a scalable and effective means to improve medication adherence in chronic disease populations [11,12]. While there has been an interest in developing health technologies such as reminder applications [13-16] or automated phone reminders for older populations [17], a review of the literature reveals that very few programs have explored using text messaging with seniors to improve medication refill adherence [18,19].

We determined at the outset that since the target users for the program were an older and/or disabled population on Medicare, it would be important to focus on usability (ie, ease of use) and simplicity (ie, design for basic feature mobile phone instead of smartphone). We used Davis' technology acceptance model (TAM) [20] as a guide to predict and optimize user acceptance of our solution as a viable and dynamic channel for interactive communication [21]. Therefore, our content strategy focused on usefulness and ease of use by providing simple instructions for authentication and task completion [22].

Objectives

The program objectives were to assess the impact of an interactive and easy-to-use text messaging solution on medication refills and pharmacy operations and efficiencies. The target population consisted of partially adherent and nonadherent Medicare patients of a large integrated health plan (Kaiser Permanente Southern California, or KP) with 1 or more chronic diseases.

Our hypothesis was that patients receiving text message refill reminders (text messaging group) in addition to existing outreach would have a higher medication refill rate compared to the group that did not receive text messages (nontext messaging group).

Methods

Participants

The program began on December 7, 2016. All patients were Medicare members of KP with 1 or more chronic conditions (diabetes, hypertension, and/or high cholesterol). Patients in this program would be refilling 1 or more of the following 3 classes of drugs: oral diabetes medications, blood pressure medications (renin-angiotensin system antagonists), and statins.

There were approximately 5000 to 14,000 patients each week on the list who required pharmacy follow-up. These patients were pulled from 3 separate KP lists: (1) New Start: patients who filled their medication the first time in the calendar year and had a day supply remaining (DSR) of 0 to 30 days, (2) Near Goal: patients whose DSR ranged from –7 to 7 days and PDC ranged from 78% to 85%, and (3) Nonadherent: patients who had 2 fills within the calendar year and need to refill their medication by a specific date (Nonadherent date) in order to have a chance to improve their PDC to 80% or higher. The Nonadherent list patients were messaged in month 1 (December 2016) only.

Patients were divided into 2 groups:

1. Text message group (12,272/88,340, 13.89%): those who had opted in to receive text messages (as recorded within the health system's electronic medical records [EMR]) and were on the weekly list for pharmacy follow-up (1000 to 4000 patients per week). These patients received text messages reminding them to refill their prescriptions. This group consisted of 12,272 patients who had opted in to receive text messages and did in fact receive text messages

over the course of the program. Table 1 provides age and race/ethnicity breakdowns for this group.

2. Nontext message group (76,068/88,340, 86.11%): those who had not opted in to receive text messages or there was no indication of an opt-in (as recorded within the health system's EMR) and were on the weekly list for pharmacy follow-up (4000 to 10,000 patients per week). This group consisted of 76,068 patients who did not receive text messages over the course of the program.

The text messaging group was one-fifth the size of the nontext messaging group because we were targeting only those Medicare patients who had opted in to receive text messages from KP. Both groups also received usual care which included phone calls and/or robo-calls reminding them to refill their prescriptions.

The Kaiser Permanente Southern California Institutional Review Board determined that this program did not involve human subject research and review was not necessary.

Procedure

Solution Overview

The mPulse Mobile platform delivers text messages to patients and members on behalf of health care companies. The platform consists of several components that together enable companies to interactively engage with their end-users about appointments, refills, gaps in care, or other health-related topics. Patients in the text messaging group received a refill reminder dialogue that consisted of a series of messages. All messages were written at a 6th grade readability level. The first message was a greeting reminding them that they were due for a refill. They were then prompted to enter their date of birth to authenticate and view their refill order (Figure 1).

Upon confirmation of the order by the patient, the KP pharmacy received a notification, and a KP pharmacist would process the refill and use the mPulse Engagement Console to inform the patient when it would be ready for pickup. Patients who did not respond to the initial message in the dialogue would receive follow-up reminders 2 hours later and again 24 hours later. They would then be moved through the same process (authentication, confirming refill order, etc). After confirmation of the order, there was no further communication with the patient. However, a small subset of patients was messaged again in a later dialogue because they failed to refill again the following month. A more detailed view of the dialogues and the process is provided in Multimedia Appendix 1.

Table 1. Characteristics of text messaging group.

Characteristic	Value, %
Age, years	
Under 65	13.2
65-70	39.7
70-75	24.1
75-85	18.9
Over 85	4.1
Race/ethnicity	
White	41.6
Hispanic/Latino	30.0
Black/African American	14.7
Asian/native Hawaiian	10.9
Unknown	2.75

Figure 1. Overview of message flow within refill dialogue.

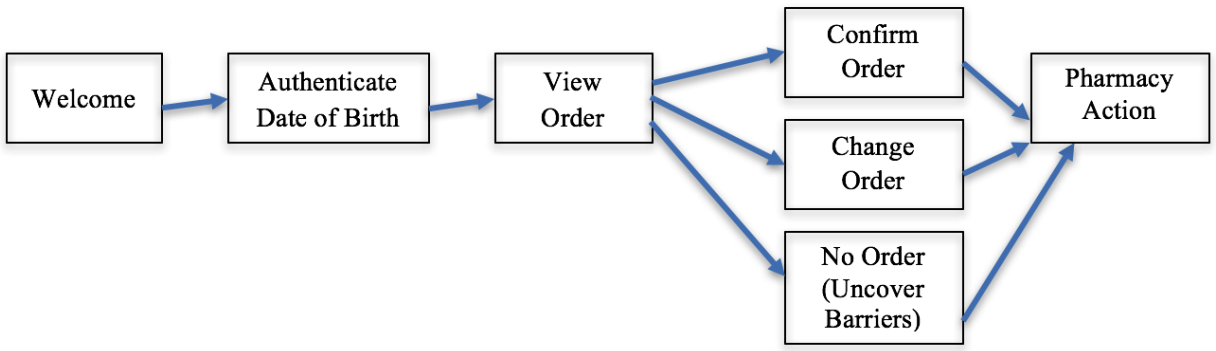
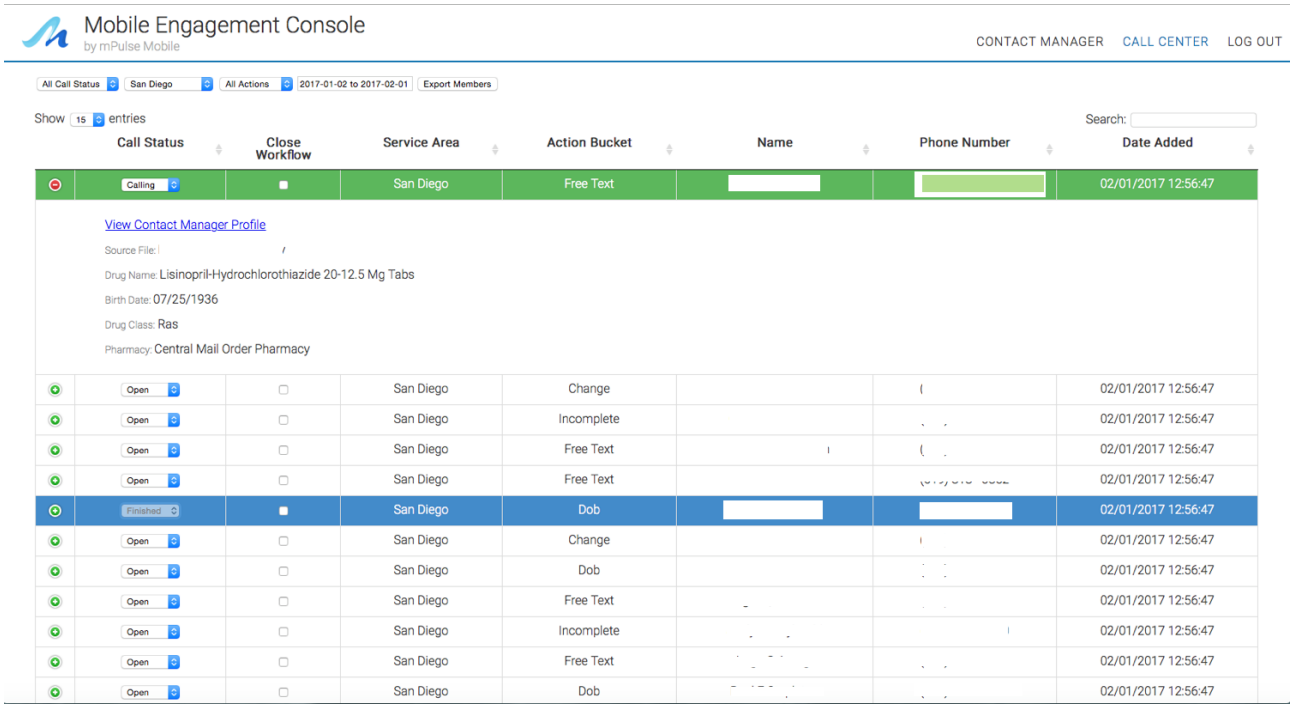


Figure 2. Engagement Console used to process refill requests and address other concerns via text.



Patients could move through the dialogue and authenticate their date of birth, complete a refill, ask for help, share reasons why they had not refilled already, or choose to opt out by using numeric or textual responses on their phone. The simplicity of the process allowed older users, who might also be more likely to have mobile phones instead of smartphones, to express their preferences and complete the process very easily.

If patients responded that they were experiencing side effects; did not believe the medication was helping them; wanted to change their medication, dose, or pharmacy; or had other concerns that might require follow-up, mPulse Mobile sent a daily list of members with pending questions or issues to the KP pharmacy for follow-up.

Dialogue Initiation

Refill dialogues were initiated at 10 am every Wednesday and Thursday to allow for a reasonable time frame within which patients could respond. Patients who texted STOP or 7867 (easier option for those without smartphones) would be opted out from the campaign and would not receive any further messages. Dialogues included tailored information to customize the message content (eg, name, date of birth, drug, pharmacy).

Patient information such as phone number, drug names, gender, name, mobile opt-in, level of adherence, and date of birth was used in 2 ways: to tailor message content for patients and initiate reminder dialogues to patients based on exclusion and combination logic. This logic helped avoid duplication and over-messaging (eg, member on multiple lists or multiple drugs would still receive a single dialogue). Patient information was provided weekly from the integrated health system and was used to perform dialogue assignments every week.

Refill Requests and Processing

Refill requests, questions, and concerns were handled by the pharmacy staff with a total of 8 staff members being trained on how to use the Engagement Console. To process refill requests or other concerns, staff would log on to a personalized view of the Engagement Console (based on their assigned medical center) and would be able to process any refill requests and other follow-up actions by initiating text messages directly to patients. They were provided with a list each week containing action buckets such as “refill requests,” “change requests,” “date of birth authentication failed or incomplete,” “help requests,” “concerns about side effects,” and “other free text responses” and prioritized their handling of these action items. Figure 2 provides a view of the Engagement Console. Additional images of the Engagement Console are provided in Multimedia Appendix 1.

Initially, processing refill requests via the Engagement Console took an average of 10 to 15 minutes. After the first week, time needed to process refill requests via the Engagement Console dropped to about 5 to 10 minutes per patient.

Results

Refill Request Rate

Our primary process measure was the number of refill requests. Of 13,195 dialogues initiated, we received a total of 2405 text messages requesting refills (Table 2). These requests were then processed by the pharmacy team and tracked separately.

Table 3 shows the number of patients targeted and the percentage who refilled by patient list. The refill request rate was highest for the Near Goal patients (1581/8206, 19.27%).

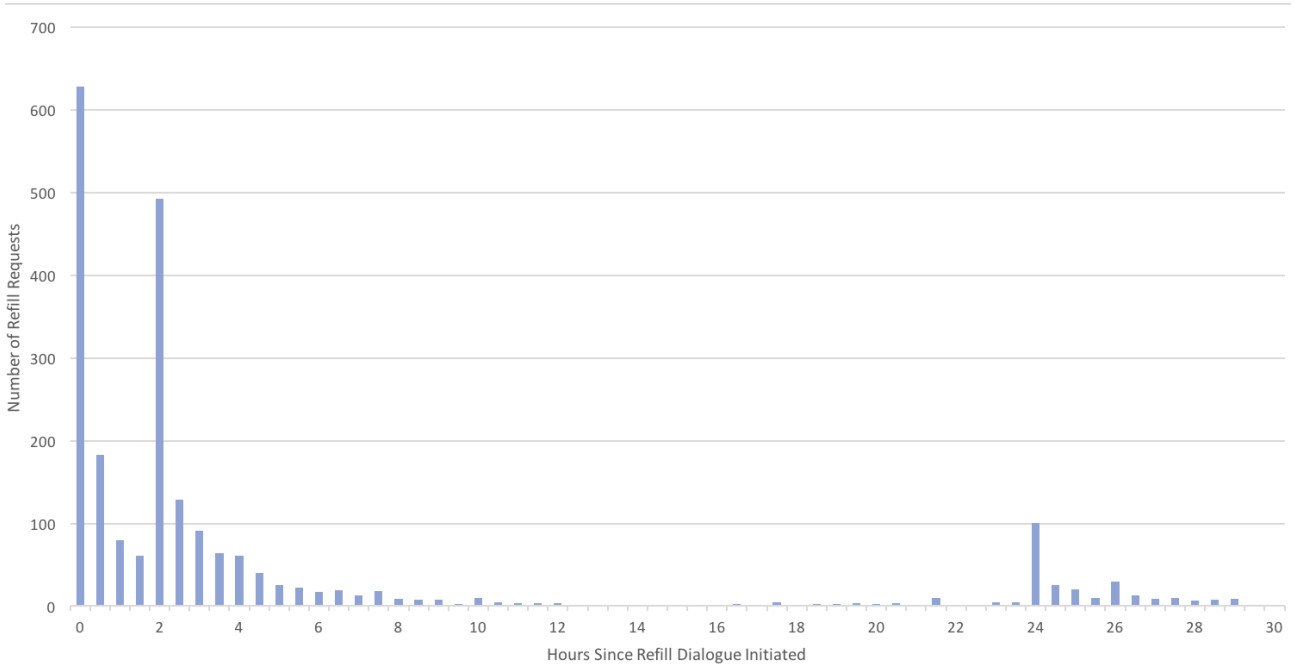
Table 2. Refill request rate for text message group by month.

Month	Refill dialogues, n	Refill requests, n	Refill request rate, %
Month 1	6776	1140	16.82
Month 2	3190	647	20.28
Month 3	3229	618	19.14
3-month total	13,195	2405	18.23

Table 3. Refill request rate for text message group by adherence level.

Adherence level	Refill dialogues, n	Refill requests, n	Refill request rate, %
Near Goal	8206	1592	19.40
New Start	748	120	16.04
Nonadherent	4241	693	16.34
Group total	13,195	2405	18.23

Figure 3. Refills requests by hour from initial reminder.



Time to Request Refill

Of those who requested a refill, 37.33% (898/2405) did so within 2 hours of receiving the initial reminder, an additional 48.61% (1169/2405) refilled within 24 hours (after also receiving the 2-hour reminder), and the remaining 14.05% (338/2405) refilled after receiving the 24-hour reminder. As displayed in Figure 3, there are spikes in refill activity immediately after the initial message (0), after the 2-hour reminder (2), and the 24-hour reminder (24). On average, members engaged within 24 minutes of getting the first message, and the median time to move through the refill process after engaging was less than 2 minutes.

Refill reminder dialogues were initiated between 10 am and noon on Wednesdays and Thursdays to allow for a reasonable time frame within which patients could respond. The bulk of refill requests (2210/2405, 91.89%) were made between the hours of 10 am and 6 pm (Figure 4). A majority of responses were received within the first 4 hours, and 81.12% (1951/2405) of responses were received within the first 8 hours.

We tracked refill request processing by pharmacy staff (total of 8 KP staff members) and noted that they collectively processed about 40 to 50 refills in an hour by the end of the first month of the program. Anecdotal feedback from KP pharmacy staff suggests that this improvement in processing refill requests has allowed them to double monthly refills.

Refills Processed

Our primary outcome measure was the number of refills that could be attributed to the text messaging. We were measuring the incremental effect of text messages (in addition to usual care) in increasing medication refills.

Figure 4. Percentage of refills requests by time of day.

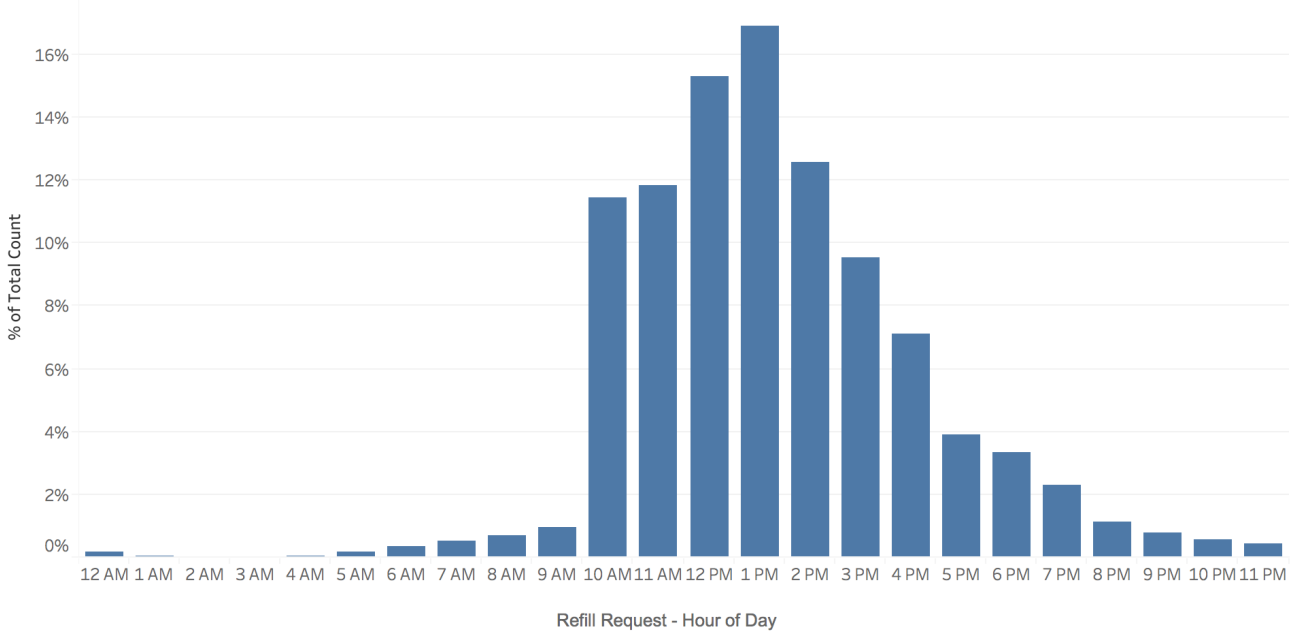


Table 4. Differences in refill rates between the text message and nontext message groups.

Month	Text message group refill rate, %	Nontext message group refill rate, %	Difference in refill rates Percentage points	*P* value
Month 1	35.73	23.49	12.24	.001
Month 2	52.55	39.10	13.45	.001
Month 3	54.05	43.23	10.82	.001
3-month total	44.08	30.01	14.07	.001

In the text message group, 12,272 patients received refill reminders via text (in addition to other outreach) over the 3-month program, and 5410 (44.08%) of these patients refilled their medication. The nontext message group of 76,068 patients received flyers and other outreach but no text reminders, and 22,826 (30.01%) of these patients completed medication refills (Table 4). The text message group refill rates were much higher than the nontext message group rates, and the 14.07 percentage point difference in refill rates between the 2 groups was statistically significant ($P<.001$).

Opt-Out Rates

The opt-out rate can be calculated in multiple ways and ranges from 1.02% to 5.09% depending on the calculation used. A total of 505 patients opted out over the course of the 3-month program. We have provided 3 different calculations in Table 5.

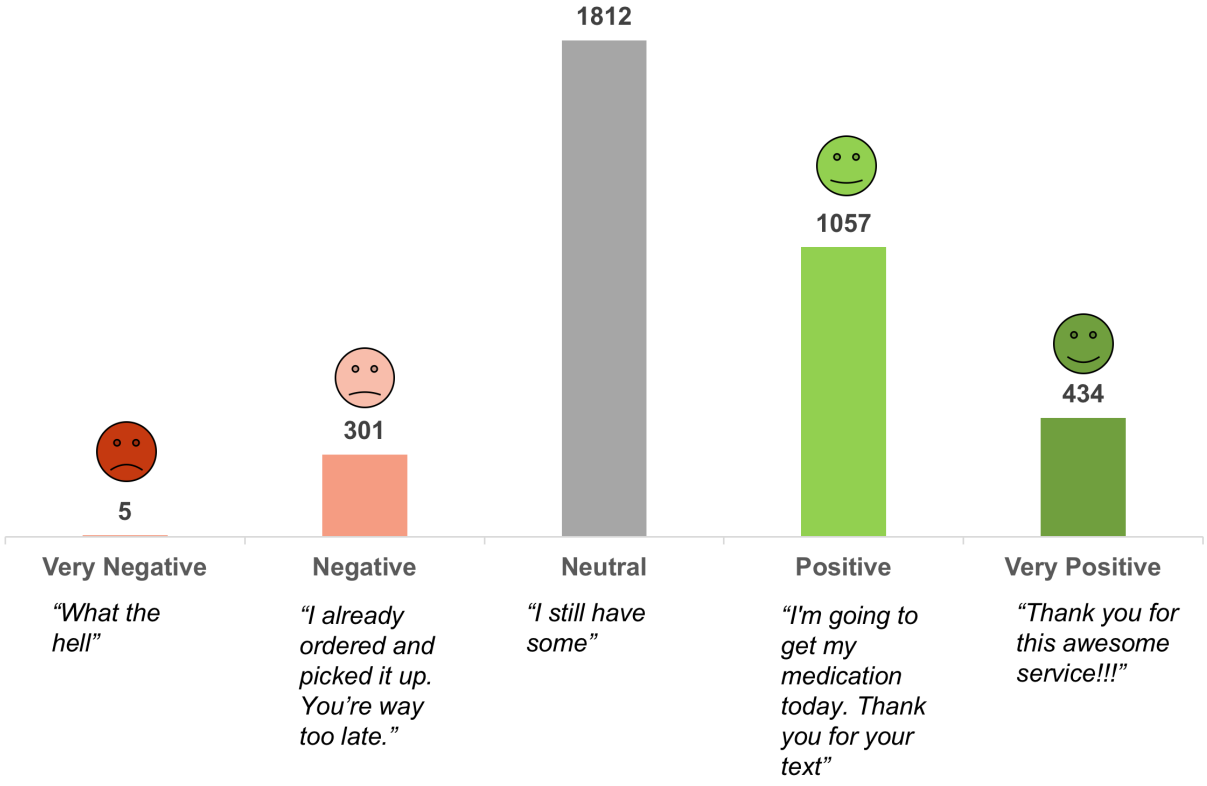
Here are the 3 different methods for calculating opt-out rates and rationale for each:

- Message level: This opt-out metric is calculated by dividing the number of members who opted out by the number of messages all members received. This measure helps us understand how long a member has stayed based on total volume of messages.
- Dialogue level: This opt-out metric is calculated by dividing the number of members who opted out by the number of dialogues all members received. This looks at the entire engagement in order to understand how well members received the program.
- Member level: This is the most common opt-out metric and is simply defined by dividing the number of members who opted out by the number of members at the beginning of the program. While this metric is useful, it does not factor in either program length or message volume and therefore presents a more coarse-grained view of member engagement and program value.

Table 5. Opt-out rates.

Approach for calculating opt-outs	Basis, n	Opt-out rate, %
Message level, messages	49,590	1.02
Dialogue level, dialogues	13,195	3.83
Member level, patients	9920	5.09

Figure 5. Sentiment in patient responses.



Measuring User Experience

We analyzed patient free text responses to understand their experience and be more responsive. To do this, we used natural language processing to extract polarity, valence, and sentiment (very positive, positive, neutral, negative, very negative). For example, "Leave me alone" has a very different emotional tone than "Thanks so much for the reminder!" As shown in Figure 5, the largest subgroup of responses was neutral (1812/3609, 50.21%), followed by positive (1057/3609, 29.28%), very positive (434/3609, 12.03%), negative (301/3609, 8.34%), and very negative (5/3609, 0.14%).

Ease of Use Survey Results

Another way in which we captured user experience was by asking patients directly. Starting in month 2, when patients completed a refill request, they received a confirmation message and were asked "Was this refill process easy to use?"

This question was intended to measure whether the TAM model's "ease of use" consideration had been successfully embedded in the refill dialogue solution. In designing for usability, we had prioritized the importance of creating a text-based refill dialogue that was easy to use, easy to learn, did not cause users to generate many errors, and was helpful to users. Over 70.02% (890/1271) of those who were presented with the survey question completed it. Of the 890 unique patients who completed the survey, 850 (95.51%) responded "Yes," and 40 (4.49%) responded "No."

Discussion

Principal Findings

We studied the value of an interactive text message refill solution with a chronically ill and partially adherent or nonadherent Medicare population and observed a difference of 14.07 percentage points in refill rates between the text message group and comparison group (P<.001).

It is worth noting that patients in the texting group engaged at a much higher rate than predicted. We had estimated that the patient response rate would be between 10% and 20%, including stop requests, help requests, date of birth authentication attempts (successful and failed), refill requests, change requests, reasons for not refilling, and other free text responses. Our target refill request rate was 5% to 7% since we were messaging an older patient population. At the same time, we hoped that the ease of use of the refill dialogue might draw in more patients and nudge them toward completing their refill requests.

The program results far exceeded our expectations. Throughout the 3-month program, the response rate was around 37%, and the 3-month average refill request rate was 18%. We had also expected that since this was an older patient population the response time span might be stretched out a little longer, but this was not the case with over 80% of refill requests received within 8 hours of the initial reminder.

We used rules and basic natural language processing to improve recognition and handling of member responses over the course of the program, cut down unprocessed free text responses from

26% to under 16%, and reduced manual handling by pharmacy staff.

Overall patient feedback was very favorable and sentiment analysis of the responses revealed that patients were 5 times more likely to express positive sentiment than negative sentiment. Finally, almost 96% of the patients who completed refills via text message indicated that the solution was easy to use, and this strongly validated the TAM model and usability considerations that guided our design of the refill dialogues.

Although a cost-effectiveness analysis was not performed, interactive text messaging is inexpensive compared to manual calls or robo-calls. Finally, the high response rates and highly positive sentiment indicates improved patient engagement with their health care provider.

Future Considerations

Our program incorporated basic demographic and psychographic data but did not tailor workflows based on the social determinants of health (ie, the conditions where people live, learn, work, and play and how these conditions affect their health risks and outcomes). This is an approach we plan to expand and implement in future programs. For example, how does living in a remote or rural area with no transportation impact refill behavior? How is income associated with refill rates? What about language and cultural barriers? This was a racially and ethnically diverse patient population. While the 3-month program used only English dialogues, the next phase would explore whether Spanish-speaking patients should be targeted differently and should also consider cultural and language barriers. We would also like to tailor content based on health literacy levels.

In future programs, we hope to combine demographic data (zip code, gender, age) with psychographic measures (adherence levels, past refill behavior, barriers to adherence, self-efficacy, stage of change, health beliefs) to develop a deeper understanding of the population being targeted to uncover health disparities and drive positive and sustained behavior change.

As we expand the program to other Kaiser Permanente regions, we expect to rely more heavily on machine learning–based natural language processing to improve recognition accuracy. Our machine learning–based natural language processing classifies a member's response into most commonly occurring categories which, in turn, triggers appropriate actions. We use a model that is topic-specific and trained on data that is based on a combination of responses received within the program and gathered through other sources. While we also rely on human intelligence to validate and handle outliers and unexpected responses, our goal is to reduce manual processing of member queries and responses to less than 5% in future programs.

Conclusion

Findings suggest that partially adherent or nonadherent Medicare patients who receive interactive text message refill reminders have significantly higher medication refill rates compared to similar patients who do not receive text message refill reminders. The program results demonstrate that this incremental value of interactive text messages increased refill rates by 14.07 percentage points in Medicare patients.

Results of the program include increased refill rates and high levels of patient engagement. These results should provide insights for developing similar models that represent an elevated standard of care within patient management.

Acknowledgments

We would like to thank the following individuals for their contributions to this project: Nico Arcino and Joyce Pedersen for initial development of project concept and scope; Allen Sarkisyan, Dustin Lo, and Ali Anari for design and development of data translation, loading, and presentation interfaces; Justine Espinoza and Nikta Moradi for overall project support; Chao Meng and Tony Askar for writing the underlying code for the dialogue engine within the mPulse Mobile platform; Matthew Betz for review and editing; and Paige Mantel for proofreading the manuscript.

Conflicts of Interest

RBP, EF, MK, and RSP are current employees of mPulse Mobile, Inc, which is a vendor for Kaiser Permanente.

Multimedia Appendix 1

Automated dialogues.

[PDF File (Adobe PDF File), 2MB - mhealth_v6i1e30_app1.pdf]

References

1. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
2. Mahoney J, Ansell B, Fleming W, Butterworth S. The unhidden cost of noncompliance. J Manag Care Pharm 2008;14:S1-S29.
3. Iuga AO, McGuire MJ. Adherence and health care costs. Risk Manag Healthc Policy 2014;7:35-44 [FREE Full text] [doi: 10.2147/RMHP.S19801] [Medline: 24591853]

4. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 4;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
5. Latif S, McNicoll L. Medication and non-adherence in the older adult. Med Health R I 2009 Dec;92(12):418-419. [Medline: 20066833]
6. Leslie RS, Tirado B, Patel BV, Rein PJ. Evaluation of an integrated adherence program aimed to increase Medicare Part D star rating measures. J Manag Care Spec Pharm 2014 Dec;20(12):1193-1203 [FREE Full text] [doi: 10.18553/jmcp.2014.20.12.1193] [Medline: 25443513]
7. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 2;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
8. Dailey J. Consumer survey on patient engagement.: Accenture; 2013. URL: https://www.accenture.com/us-en/~/media/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_11/Accenture-Silver-Surfers-Catching-eHealth-Wave-Infographic [accessed 2018-01-16] [WebCite Cache ID 6wWuXLF0B]
9. Anderson M. Technology device ownership: 2015. Washington: Pew Internet and American Life Project URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/10/PI_2015-10-29_device-ownership_FINAL.pdf [accessed 2018-01-16] [WebCite Cache ID 6wWv6WkR1]
10. Rainie L. Digital divides. Washington: Pew Research Center URL: http://www.pewinternet.org/2016/07/14/digital-divides-2016/ [accessed 2018-01-16] [WebCite Cache ID 6wWv2CJrr]
11. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
12. Pandey A, Krumme A, Patel T, Choudhry N. The impact of text messaging on medication adherence and exercise among postmyocardial infarction patients: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Aug 03;5(8):e110 [FREE Full text] [doi: 10.2196/mhealth.7144] [Medline: 28778843]
13. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
14. Ferreira F, Almeida N, Rosa AF, Oliveira A, Casimiro J, Silva S, et al. Elderly centered design for interaction and the case of the S4S medication assistant. Procedia Computer Science 2014;27:398-408. [doi: 10.1016/j.procs.2014.02.044]
15. Dalgaard L, Gronvall E, Verdezoto N. MediFrame: a tablet application to plan, inform, remind and sustain older adults' medication intake. 2013 Presented at: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI); 2013; Philadelphia p. 36-45.
16. Silva J, Mouttham A, El Saddik A. UbiMeds: a mobile application to improve accessibility and support medication adherence. 2009 Presented at: Proceedings of the 1st ACM SIGMM International Workshop on Media Studies and Implementations that Help Improving Access to Disabled Users; 2009; Beijing p. 71-78. [doi: 10.1145/1631097.1631109]
17. Vollmer WM, Owen-Smith AA, Tom JO, Laws R, Ditmer DG, Smith DH, et al. Improving adherence to cardiovascular disease medications with information technology. Am J Manag Care 2014 Nov;20(11 Spec No. 17):SP502-SP510 [FREE Full text] [Medline: 25811824]
18. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
19. Penno E, Gauld R. Change, connectivity, and challenge: exploring the role of health technology in shaping health care for aging populations in Asia Pacific. Health Systems & Reform 2017 Aug 08;3(3):224-235. [doi: 10.1080/23288604.2017.1340927]
20. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
21. King WR, He J. A meta-analysis of the technology acceptance model. Information & Management 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
22. Day M, Demiris G, Oliver DP, Courtney K, Hensel B. Exploring underutilization of videophones in hospice settings. Telemed J E Health 2007 Feb;13(1):25-31. [doi: 10.1089/tmj.2006.0023] [Medline: 17309351]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
DSR: day supply remaining
EMR: electronic medical records
KP: Kaiser Permanente Southern California
PDC: proportion of days covered
SMS: short message service

TAM: technology acceptance model

Edited by G Eysenbach; submitted 08.09.17; peer-reviewed by J Thakkar, J Redfern; comments to author 29.09.17; revised version received 20.10.17; accepted 08.01.18; published 30.01.18.

Please cite as:
Brar Prayaga R, Jeong EW, Feger E, Noble HK, Kmiec M, Prayaga RS
Improving Refill Adherence in Medicare Patients With Tailored and Interactive Mobile Text Messaging: Pilot Study
JMIR Mhealth Uhealth 2018;6(1):e30
URL: http://mhealth.jmir.org/2018/1/e30/
doi:10.2196/mhealth.8930
PMID:29382623

Original Paper

Prevalence of Health App Use Among Older Adults in Germany: National Survey

Peter Rasche[1], MSc; Matthias Wille[1], Dr Phil, Dipl-Psych; Christina Bröhl[1], MSc; Sabine Theis[1], MSc; Katharina Schäfer[1], MA; Matthias Knobe[2], Priv-Doz, Dr med, MME, MHBA; Alexander Mertens[1], Dr-Ing, Dr rer medic, Dipl-Inform

[1]Department of Mechanical Engineering, Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany
[2]Department of Orthopaedic Trauma, University of Aachen Medical Center, RWTH Aachen University, Aachen, Germany

Corresponding Author:
Peter Rasche, MSc
Department of Mechanical Engineering
Institute of Industrial Engineering and Ergonomics
RWTH Aachen University
Bergdriesch 27
Aachen, 52062
Germany
Phone: 49 2418099477
Email: p.rasche@iaw.rwth-aachen.de

Abstract

Background: Health apps are increasingly becoming an integral part of health care. Especially in older adults, the self-management of chronic diseases by health apps might become an integral part of health care services.

Objective: The aim of this explorative study was to investigate the prevalence of health app use and related demographic factors, as well as health status among older adults in Germany.

Methods: A nationwide postal survey was conducted. Of the 5000 individuals contacted, a total of 576 participants completed this survey. On the basis of their self-indicated assignment to one of the three predefined user groups (health app users, general app users, and nonusers of apps), participants answered various questions regarding app and health app use, including frequency of use and number of installed apps, demographic factors, and health status.

Results: In total, 16.5% (95/576) used health apps, whereas 37.5% (216/576) indicated only using general apps, and 46.0% (265/576) reported using no apps at all. The number of installed health apps was most frequently reported as between 1 and 5 apps per participant, which were usually used on a weekly basis. The most frequently cited type of health apps were exercise-related ones. Individuals using health apps were found to be younger (MeanmHealth 66.6, SD 4.7) and to have a higher level of technical readiness compared with general app users and nonusers of apps (adjusted odds ratio, AOR=4.02 [95% CI 2.23-7.25] for technical readiness, and AOR=0.905 [95% CI 0.85-0.97] for age). The most frequently mentioned sources of information about apps within the group of health and general app users were family and friends. Identified barriers against the use of health apps were a lack of trust, data privacy concerns, and fear of misdiagnosis.

Conclusions: Health apps are already used by older adults in Germany. The main type of apps used are exercise-related ones. Barriers to and incentives for the use of health apps and associations with health status and users' demographics were revealed.

(JMIR Mhealth Uhealth 2018;6(1):e26) doi:10.2196/mhealth.8619

KEYWORDS

telemedicine; Germany; mobile applications; smartphone; aged

Introduction

The number of households with individuals older than 75 years will double by 2050. Fifty percent of the population of Germany will then be older than 50 years and 12% older than 80 years [1]. In this context, aging is often associated with a decline in quality of life (QoL) because of factors such as limited mobility and autonomy and the appearance of various chronic diseases [2]. By providing older adults with suitable health apps, they might be enabled to monitor and manage health in an

independent and self-determined manor [3-7]. Smartphones have a large number of sensors able to measure and track vital parameters as well as other health-related data [3]. Health apps analyze and process these data and could therefore provide integral support to health care services in the future [3,6,8]. There are numerous examples of such health apps that cover diseases such as diabetes or other chronic conditions as well as topics such as fall risk. To date, only a small number of health apps have been introduced into the health care market [9-12].

As health apps are a relatively new phenomenon, it is not yet known to what extent health apps are being accepted and used by older adults for personal health care. Initial studies into the use of health apps among the general population have been carried out in the United States and Hong Kong [13,14]. In the United States, 19% of smartphone owners had used at least one health app in 2012 already [14]. The study did not, however, measure the level of use among older adults within the sample. A more recent study revealed a use rate in 2015 of more than 50% within a sample of younger and older adults in the United States [15]. A different study in Hong Kong identified a use rate of about 20% among this population in 2016 [13]. The most frequently mentioned health apps in all these studies were exercise-related. Nevertheless, the question remains whether older adults are willing to integrate such communication-technology–based health care products into their daily lives. For them, this means that they would have to adapt to the new technology, especially as they are "digital immigrants" who might not be engaged with smartphones and related products [16].

To the best of our knowledge, we were the first to investigate this question by evaluating whether older adults in Germany have any health apps installed on their mobile devices, how many apps they have installed, and how often they use them.

Aim of This Study

In this study, we investigated the prevalence of health apps among older adults in Germany, incentives for and barriers to their use, as well as which sources users rely on to gain information about health apps. We performed a postal survey questioning possession and use of health apps related to demographic factors and health status as measured by a health literacy scale and history of chronic diseases [17].

Research Questions

In summary, our main research questions were as follows:

- Do older adults in Germany use health apps?
- What types of health apps are currently being used by older adults?
- What factors are associated to health app use?
- What sources do older adults rely on to retrieve information about apps?

Methods

Design

A postal survey was designed to investigate the aforementioned research questions. The survey was designed in German and provided for German-speaking older adults. A postal survey was used as it is a cost-effective way to reach older adults in a short period of time without any limitations on physical space [18,19].

On the basis of the research questions, the main purpose of the survey was to collect data about three different user groups that we would like to compare. The three predefined user groups we wanted to identify and compare were older adults who use health apps ("health app users"), those who use general apps but not health apps ("general app users"), and those who had never used an app ("nonusers of apps"). To identify these three groups, users were asked to state to which of them they belong.

Characterizing Participants

To characterize users of health and general apps, more detailed questions about the number of installed apps and frequency of use, as well as health status, were asked.

Measuring Technology Readiness

Technology readiness was included as it might influence the use of modern information and communication technology, as well as the engagement with these products [20]. It is calculated based on 12 standardized items which are rated on a 5-point Likert scale (1=not correct, 5=fully correct). For positively formulated objects, the scale is converted so that a high point value corresponds with high technology readiness. Subsequently, the final score is calculated by mean value over all 12 items. The score therefore ranges between 1 point and 5 points [20].

Measuring Computer Literacy

Computer literacy was included as it identifies whether an individual has high or low experience in using computers [21]. Computers have been available for a much longer period of time and are more widely used than smartphones. We therefore wanted to determine whether high experience with computers might influence the use of health apps, as such individuals might have lower barriers to using this technology. The computer literacy scale questionnaire was used in a shorted version with 20 items. On the basis of a dichotomous coding of the responses (false=0, correct= 1), the score was calculated as over the sum of all items. The final score, therefore, ranges from 0 points up to 20 points, whereby a high score indicates high computer literacy [21].

Measuring Health Status

Health apps are a part of the field of consumer health, so it is interesting to investigate whether an association between health app use and health status exists [6]. Shen et al identified an association between history of chronic disease and health app use among the general population of Hong Kong [13]. We wanted to know whether this is the same for older adults in Germany. Participants were asked whether they suffer from chronic diseases and, if so, which type. Additionally, the health competency of participants was measured using an adapted version of the European Health Literacy Survey-47 (EU HLS-47), with 15 items [17]. Corresponding statements were evaluated on a 5-point Likert scale (1=not correct, 5=fully correct). Subsequently, the point values were converted into dichotomous values and summed [22]. Therefore, the final score

ranges between 0 points and 15 points, whereby a high score indicates high health competency [17].

Investigating Use of Health Apps

The main aim of this study was to determine whether older adults in Germany use health apps. Therefore, all participants were asked whether they use health apps or not. If participants indicated using health apps, they were sorted into the group of health app users. Besides the simple question of whether participants use health apps or not, the number of health apps installed on participants' smartphones was also inquired. In addition, participants were asked how frequently they use their health apps and which type of health app they use in particular. As these figures alone are not meaningful, participants were also asked about the number of general apps installed and their usage frequency.

Investigating Acceptance of Health Apps

A second goal of this study was to investigate the barriers to the acceptance of health apps. Both health app users and general app users were surveyed. The group of nonusers of apps was not considered in this question because the lack of experience with apps was assumed to be too influential. Both other groups (health app users and general app users) were asked what reasons they would personally argue against using health apps. In addition to a free text field, further answers were given. These reasons include privacy concerns, lack of self-confidence, fear of misdiagnosis, poor usability, and lack of confidence in the app. Multiple answers were possible for this question.

Identifying Sources of Information About Apps

The third goal of this study was to identify different sources of information older adults use to find information about apps. Here too, only health app users and general app users were surveyed. The participants were able to select from different predefined answers the ones that fit them best. Available sources of information were family and friends, Internet, app store, magazines, television, and experts, with experts the least frequently cited source of information.

Data Collection

Data were collected from July 2016 to September 2016. The questionnaire was printed and sent by postal mail. The survey was introduced as a study examining the effects of modern digital technology on the German health care system (see Multimedia Appendix 1).

All of the participants were informed about the duration of the survey, data storage, and the leading investigator. Each participant decided to take part in this survey voluntarily by sending back the completed questionnaire in a postage-paid envelope. No incentives were offered for participation.

The survey was tested properly by five independent examiners with regard to wording. The paper-based questionnaire included 109 items distributed over 15 different pages.

Recruitment

The survey was addressed to 5000 older adults retrieved from the general German population by the external service provider Deutsche Post AG. Inclusion criteria for the selection by Deutsche Post AG were older than 60 years and residence in Germany. No exclusion criteria or screening questionnaires were applied.

All questionnaires were sent by post. The sampling procedure was nonprobabilistic, and respondents were self-selected based on their voluntary willingness to participate. This method of recruitment was chosen as the probability that participants with no or limited experience in the topic of apps would be reached by a Web-based survey is quite low [19]. Bech and Kristensen investigated differences in response rates regarding a potential future design of nursing homes. Their study revealed a higher response rate for a postal than a Web-based survey among older adults [23]. Hence, a postal survey was conducted in this study. Furthermore, this postal survey is an observational study targeting participants who use or do not use general and health apps. Recruitment via postal mailing, therefore, seems to be a suitable and cost-effective approach [18,19].

In total, N=5000 unique individuals were contacted by postal mail. However, only n=576 contacted older adults participated in the survey and completed the questionnaire. The participation rate is thus 11.52% (576/5000). Due to the method of data collection, the completion rate is unknown. On the basis of pretests, it can be stated that the average time required to complete the survey was around 45 min.

Statistical Analysis

Data were analyzed using Statistical Package for the Social Sciences (SPSS) version 22 (IBM Corp). The associations of age, sex, education, technical readiness, computer literacy, health competence, and history of diagnosed chronic disease with health app possession were analyzed by logistic regression in a model with these variables mutually adjusted. The associations of age, sex, education, technical readiness, computer literacy, health competence, and multimorbidity (defined as the number of chronic diseases participants suffer from) with the different reasons for not accepting health apps and preferred sources of information were analyzed by logistic regression in separate models with these variables mutually adjusted. To compare general and health app users, we also calculated *t* tests for independent samples and chi-square statistics both at a significance level of .05.

Ethics Statement

The ethics committee at RWTH Aachen Faculty of Medicine authorized this study and its ethical and legal implications in its statement EK236/16.

Results

Participants

In total, 576 individuals took part in this study. Mean age was 69.17 years (SD 5.76). Of the 576 participants, 280 (48.7%) were female. Some 10.1% (57/576) of the participants achieved primary education level, 57.3% (330/576) achieved secondary education level, and 32.5% (187/576) achieved tertiary education level. A total of 286 out of 576 participants (49.6%) use a smartphone and 132 out of these 286 smartphone users (46.2%)

also use a tablet. Just 33 out of 576 participants (5.7%) use a tablet without an additional smartphone.

Depending on participants' statement regarding health or general app use, they were divided into three user groups: users of health apps (n=95), users of general apps (n=216), and nonusers of apps (n=265). Table 1 shows detailed sociodemographic information for participants as sorted into the three different groups.

On average, users in all three groups reported at least one chronic disease per participant ($Mean_{mHealth}$ 1.3 [SD 1.3], $Mean_{general}$ 1.1 [SD 1.1], and $Mean_{nonusers}$ 1.1 [SD 1.1]). The reported number of chronic diseases did not differ significantly between the three groups, $F_{1,309}$=0.08, P=.78. Table 2 provides an overview of the different types of chronic diseases reported. All groups mentioned hypertension most frequently. Back pain, arthrosis, and diabetes were also frequently mentioned chronic diseases within all three groups (see Table 2). The only significant differences between the three user groups occurred for osteoporosis and back pain; health users did not mention osteoporosis at all, χ^2_2=7.1, P=.02, and back pain was reported most frequently by nonusers, followed by health app users and then general app users, χ^2_2=8.5, P=.01.

Use of Health Apps

This section reports the findings relating to health app use. We focus especially on predictive factors to identify incentives for and barriers to getting older adults engaged with health apps.

Participants were asked how many general apps they have installed. The two groups (health app users and general app users) differ significantly, χ^2_4=24.7, P<.001. The most frequently cited number was "up to ten" for both health app users (37%, 35/95) and general app users (60%, 129/216). Additionally, participants were asked how often they use general apps. The most frequent answer within the group of health apps users was "daily" (73%, 69/95), the same as for the group of general app users (49%, 106/216). However, the two groups differ significantly for time of use, χ^2_4=17.6, P<.001. Table 3 reports the detailed frequencies.

Table 1. Participant demographics by user group.

Participant demographics	User group		
	Health app users	General app users	Nonusers of apps
Sample			
Number of participants, n	95	216	265
Age, years			
Minimum	61	60	61
Maximum	82	80	90
Mean (SD; years)	66.6 (4.7)	68.2 (5.1)	70.8 (6.1)
Gender			
Male, n (%)	54 (56.8)	117 (54.2)	124 (46.8)
Female, n (%)	41 (43.2)	98 (45.8)	141 (53.2)
Education			
No education, n (%)	0 (0.0)	0 (0.0)	1 (0.3)
Low level, n (%)	3 (3.1)	16 (7.4)	37 (14.0)
Average level, n (%)	42 (44.2)	99 (45.8)	140 (52.8)
High level, n (%)	49 (51.6)	95 (44.0)	78 (29.4)
Other, n (%)	1 (1.1)	6 (2.8)	7 (3.5)
Technical readiness			
Range (points)	2.17-5	1.67-4.83	2.17-5
Mean (SD; points)	3.7 (0.6)	3.4 (0.6)	2.9 (0.6)
Computer literacy			
Range (points)	2-18	3-18	1-18
Mean (SD; points)	27.0 (3.9)	13.5 (3.7)	10.0 (5.1)
Health competence			
Range (points)	3.5-15	5-15	3.5-15
Mean (SD; points)	11.3 (2.8)	10.8 (2.6)	10.5 (2.9)

Table 2. Mean number and types of chronic diseases reported per group (multiple answers allowed).

Number and types of chronic diseases	User group			Significance
	Health app users	General app users	Nonusers of apps	
Number of chronic diseases, mean (SD)	1.1 (1.1)	1.1 (1.1)	1.3 (1.3)	F=0.08 (1,309), P=.78
Types of chronic diseases, n (%)				
Hypertension	37 (38.9)	80 (37.0)	106 (40.0)	χ^2_2=0.4, P=.80
Back pain	20 (21.1)	38 (17.6)	76 (28.7)	χ^2_2=8.5, P=.01
Arthrosis	17 (17.9)	36 (16.7)	53 (20.0)	χ^2_2=0.9, P=.63
Diabetes	15 (15.8)	22 (10.2)	41(15.5)	χ^2_2=3.3, P=.19
Overweight	12 (12.6)	28 (13.0)	45 (17.0)	χ^2_2=1.9, P=.38
Cardiovascular disease	8 (8.4)	24 (11.1)	42 (15.8)	χ^2_2=4.3, P=.11
Respiratory disease	6 (6.3)	14 (6.5)	19 (7.2)	χ^2_2=0.1, P=.93
Osteoporosis	0 (0)	7 (3.2)	16 (6.0)	χ^2_2=7.1, P=.02

Table 3. Number of installed general apps and time of use.

Number and frequency of use	User group		Significance
	Health app users	General app users	
Number of installed general apps, n (%)			χ^2_4=24.7, P<.001
≤10	35 (36.8)	129 (59.7)	
11-20	33 (34.7)	37 (17.1)	
21-30	16 (16.8)	15 (6.9)	
31-40	4 (4.2)	2 (0.9)	
>40	3 (3.2)	4 (1.9)	
Frequency of use, n (%)			χ^2_4=17.6, P<.001
Daily	69 (72.6)	106 (49.1)	
Every 2-3 days	17 (17.9)	41 (19.0)	
Weekly	4 (4.2)	15 (6.9)	
Monthly	1 (1.0)	18 (8.3)	
Never	0 (0)	31 (14.4)	

To examine the use of health apps in detail, we asked users how many health-related apps they have installed and how often they use them. The most frequently reported number of installed apps was “1-5 health apps” (85%, 81/95), followed by “6-10 health apps” (3%, 3/95), and just one participant reported having “11-15 health apps” (1%, 1/95). Health app users stated they have up to 20 general apps installed. Asked about the frequency of use, health app users most frequently mentioned weekly use (27%, 26/95), followed by monthly (22%, 21/95) and “every 2-3 days” (16%, 15/95), with just 10 participants claiming to use health apps on a daily basis (11%, 10/95). Comparing these frequencies with frequencies of general app use reveals a significant difference, t_{88}=−13.54, P<.001. Health apps are less frequently used than general apps.

Users of health apps were also asked about the type of health apps they use (multiple answers allowed). The most frequently reported type were exercise apps (29%, 28/95), followed by apps for rating (26%, 25/95), first aid response (13%, 12/95), diabetes management (12%, 11/95), diagnosis (12%, 11/95), emergency passport (9%, 9/95), dairy (8%, 8/95), medication planner (7%, 7/95), and for contacting the physician (1%, 1/95).

Logistics regression revealed technical readiness and age to be associated with use of health apps (AOR=4.02 [95% CI 2.23-7.25] for technical readiness and AOR=0.905 [95% CI 0.85-0.97] for age). Sex, education, computer literacy, health competence, multimorbidity, and field of work were not associated with having health apps. Descriptive comparison showed that the mean age within the group of health application users is lower than within the group of general app users. The

highest mean age was found for the group of nonusers of apps. The mean values of technical readiness show that health app users have the highest level of technical readiness followed by general app users and then nonusers ($Mean_{mHealth}$ 3.7 [SD 0.6], $Mean_{general}$ 3.4 [SD 0.6], and $Mean_{nonusers}$ 2.9 [SD 0.6]).

Acceptance of Health Apps

Participants were asked for reasons decreasing subjective acceptance of health apps. Hence, a more detailed picture is drawn of the barriers older adults might have toward using health apps. Table 4 shows that a lack of trust in health apps is the major barrier to using them.

Users of general apps mentioned significantly more reasons than health app users, t_{310}=25.37, P<.001. The most frequently mentioned reason within both groups is a lack of trust, followed by data privacy concerns and fear of misdiagnosis. The two groups differ significantly in the case of poor usability, χ^2_1=4.8, P<.02 and a lack of self-confidence, χ^2_1=4.7, P=.03. Both reasons were more frequently mentioned by users of general apps than of health apps.

Analysis of association between the reasons against acceptance and demographic characteristics revealed different associations (see Table 5). A lack of trust is positively associated with technical readiness and computer literacy (AOR=1.74 [95% CI 1.18-2.58] for technical readiness and AOR=1.08 [95% CI 1.02-1.14] for computer literacy). Furthermore, data privacy concerns are positively associated with technical readiness and computer literacy (AOR=1.66 [95% CI 1.01-2.71] for technical readiness and AOR=1.11 [95% CI 1.03-1.21] for computer literacy). The fear of misdiagnosis is associated with decreasing age and increasing multimorbidity and technical readiness (AOR=0.92 [95% CI 0.85-0.99] for age, AOR=1.42 [95% CI 1.07-1.89] for multimorbidity, and AOR=1.94 [95% CI 1.01-3.73] for technical readiness). Poor usability was associated with increasing age and decreasing health competence and decreasing technical readiness (AOR=1.08 [95% CI 1.01-1.16] for age, AOR=0.8 [95% CI 0.69-0.92] for health competence, and AOR=0.41 [95% CI 0.21-0.78] for technical readiness). Finally, a lack of self-confidence was associated with decreasing health competence and technical readiness (AOR=0.66 [95% CI 0.53-.82] for health competence and AOR=0.32 [95% CI 0.13-0.77] for technical readiness).

Sources of Information About Apps

Besides use and acceptance, we also investigated where participants acquire information about health apps. Table 6 shows which sources participants rely on.

Health app users use significantly more sources of information about apps than users of general apps ($Mean_{mHealth}$ 2.55 [SD 1.24]; $Mean_{general}$ 1.66 [SD 89]), t_{310}=31.21, P<.001. Family and friends are the most trusted and preferred source of information. The two groups differ significantly regarding the use of the Internet: χ^2_1=22.4, P<.001; app store: χ^2_1=5.3, P<.05; magazines: χ^2_1=25.3, P<.001; and television: χ^2_1=9.8, P<.05 as sources of information. Users of health apps significantly more often reported the mentioned sources as relevant for them.

For the sources of information, associations with demographic characteristics were also calculated to identify influencing factors (see Table 7). Family and friends as source of information was associated with young age and high technical readiness and high computer literacy (AOR=0.96 [95% CI 0.92-0.99] for age, AOR=1.43 [95% CI 0.99-2.04] for technical readiness, and AOR=1.1 [95% CI 1.04-1.17] for computer literacy).

Table 4. Mentioned reasons decreasing subjective acceptance of health apps (multiple answers allowed).

Reasons	User group		Significance
	Health app users	General app users	
Number of reasons mentioned, mean (SD)	1.23 (0.98)	1.45 (0.95)	t_{310}=25.37, P<.001
Mentioned reasons, n (%)			
Lack of trust	62 (65.3)	157 (72.7)	χ^2_2=1.7, P=.19
Data privacy concerns	27 (28.4)	63 (29.2)	χ^2_2=0.0, P=.89
Fear of misdiagnosis	20 (21.1)	33 (15.3)	χ^2_2=1.5, P=.21
Poor usability	6 (6.3)	33 (15.3)	χ^2_2=4.8, P=.02
Lack of self-confidence	1 (1.0)	15 (6.9)	χ^2_2=4.6, P=.03
Lack of interest[a]	1 (1.0)	12 (5.6)	χ^2_2=3.7, P=.05
Pressure to perform[a]	0 (0)	2 (0.9)	χ^2_2=3.7, P=.05
Technical reasons[a]	0 (0)	2 (0.9)	χ^2_2=3.7, P=.05

[a]Answers to open-ended answer option; coded for analysis.

Table 5. Association between demographic characteristics and selected reasons for a lack of acceptance of health apps is shown.

Demographic characteristics	Reason									
	Lack of trust (n=184), AOR[a] (95% CI)	*P* value	Data privacy concerns (n=81), AOR (95% CI)	*P* value	Fear of misdiagnosis (n=45), AOR (95% CI)	*P* value	Poor usability (n=36), AOR (95% CI)	*P* value	Lack of self-confidence (n=19), AOR (95% CI)	*P* value
Age	0.97 (0.93-1.02)	.21	.98 (0.93-1.03)	.41	0.92 (0.85-0.99)	.03	1.08 (1.01-1.16)	.03	0.98 (0.89-1.08)	.69
Education	—	.70	—	.83	—	.37	—	.38	—	.77
Multimorbidity	1.19 (0.98-1.45)	.07	1.19 (0.94-1.5)	.14	1.42 (1.07-1.89)	.02	0.987 (0.7-1.4)	.94	1.32 (0.85-2.05)	.22
Health competence	0.98 (0.89-1.07)	.65	0.93 (0.83-1.04)	.20	0.93 (0.81-1.07)	.29	0.8 (0.69-0.92)	.01	0.66 (0.53-0.82)	>.001
Technical readiness	1.74 (1.18-2.58)	.01	1.66 (1.01-2.71)	.04	1.94 (1.01-3.73)	.04	0.41 (0.21-0.78)	.01	0.32 (0.13-0.77)	.01
Computer literacy	1.08 (1.02-1.14)	.01	1.11 (1.03-1.21)	.01	1.08 (0.98-1.19)	.13	1.09 (0.99-1.2)	.08	0.99 (0.88-1.13)	.91

[a]AOR: adjusted odds ratio.

Table 6. Source of information participants rely on regarding apps (multiple answers allowed).

Sources	User group		Significance
	Health app users	General app users	
Number of sources used, mean (SD)	2.55 (1.24)	1.66 (0.89)	t_{310}=31.21, *P*<.001
Type of Sources, n (%)			
Family and friends	78 (82.1)	162 (75.0)	χ^2_1=1.8, *P*=.16
Internet	52 (54.7)	58 (26.6)	χ^2_1=22.4, *P*<.001
App store	47 (49.5)	77 (35.6)	χ^2_1=5.2, *P*<.05
Magazines	40 (42.1)	34 (15.7)	χ^2_1=25.2, *P*<.001
Television	15 (15.8)	11 (5.1)	χ^2_1=9.8, *P*<.05
Experts	10 (10.5)	16 (7.4)	χ^2_1=0.8, *P*=.36

Table 7. Association between demographic characteristics and sources of information about apps.

Demographic characteristics	Source of information											
	Family and friends (n=203), AOR[a] (95% CI)	*P* value	Internet (n=96), AOR (95% CI)	*P* value	App store, AOR (95% CI)	*P* value	Magazines, AOR (95% CI)	*P* value	Television, AOR (95% CI)	*P* value	Experts, AOR (95% CI)	*P* value
Age	0.96 (0.92-0.99)	.04	0.99 (0.94-1.05)	.87	0.94 (0.89-0.99)	.04	0.97 (0.91-1.04)	.37	0.88 (0.78-0.99)	.04	0.96 (0.87-1.05)	.33
Education	—	.16	—	.38	—	.08	—	.92	—	.94	—	.99
Multimorbidity	0.9 (0.74-1.08)	.25	1.2 (0.96-1.51)	.11	0.90 (0.72-1.13)	.4	0.87 (0.66-1.15)	.33	1.24 (0.8-1.93)	.33	1.13 (0.8-1.62)	.47
Health competence	0.94 (0.86-1.02)	.13	0.93 (0.84-1.02)	.13	0.93 (0.84-1.02)	.12	0.97 (0.86-1.08)	.56	1.01 (0.83-1.23)	.89	1.08 (0.91-1.28)	.36
Technical readiness	1.43 (0.99-2.04)	.05	3.36 (2.1-5.4)	>.001	3.11 (1.95-4.97)	>.001	3.24 (1.84-5.69)	>.001	3.22 (1.26-8.19)	.01	1.95 (0.93-4.09)	.08
Computer literacy	1.1 (1.04-1.17)	.03	1.11 (1.03-1.21)	.01	1.13 (1.05-1.22)	.001	1.13 (1.03-1.25)	.01	1.09 (0.93-1.29)	.26	1.04 (0.91-1.17)	.58

[a]AOR: adjusted odds ratio.

Discussion

Principal Findings

The potential of mobile phone apps to increase quality in health care and thus the QoL of patients is currently the topic of much discussion [3,11]. Initial studies have investigated the effects of smartphone apps on diabetes and chronic disease management, as well as fall risk [9-12]. The focus of this study lies on the investigation of the prevalence of health apps among older adults in Germany.

This explorative study presents the results of a postal survey. The response rate was about 11.52% (576/5000). The sample (N=576) was divided into three groups (users of health apps (n=95), users of general apps (n=216), and nonusers of apps (n=265). Most participants (54.0%, 311/576) use on average up to ten general apps and on a daily basis. Moreover, 16.5% (95/576) of the whole sample already use at least one health app. Participants most frequently reported having “1-5 health apps” installed on their smartphone and which they use on a weekly basis. Health apps were therefore found to be less frequently used than general apps (daily basis). Asked about the type of health app used, participants mentioned exercise apps most frequently (29%, 28/95), followed by rating apps (26%, 25/95), or apps for first aid response (13%, 12/95). Age and high technical readiness were positively associated with the use of health apps. The average number of chronic diseases and health competence as measured by an adapted version of the EU HLS-47 did not differ significantly between the three groups. Participants suffered on average from at least one chronic disease, whereby the most frequently reported diseases for all three groups were hypertension and back pain. Factors such as gender or education showed no significant differences between the three groups and thus, no influence on the use of health apps. Results about barriers of health app use showed that a lack of trust, data privacy concerns, and fear of misdiagnosis are the main ones for health as well as general app users. Differences between the two groups were revealed for poor usability (χ^2_1=4.8, *P*=.02) and a lack of self-confidence (χ^2_1=4.7, *P*=.03). These barriers were more frequently mentioned by general app users. These barriers are therefore possible reasons why general app users do not engage with health apps [5]. Finally participants were asked about their preferred source of information about apps. Users of health apps reported a significantly higher number of sources than users who only use general apps ($Mean_{mHealth}$ 2.55 [SD 1.24] and $Mean_{general}$ 1.66 [SD 0.89]). The most frequent source participants rely on is family and friends. Significant differences were revealed for sources such as Internet, app store, magazines, and television, which were more frequently mentioned by health app users than general app users. Experts were the least frequently mentioned source of information. Therefore, targeting older adults’ family and friends as motivators for health app use seems to be a promising approach.

Our study adds to existing research by reporting associated factors of health app use, frequency of use, as well as source of information regarding apps for older adults in Germany. In contrast to findings in previous national surveys in the United States and Hong Kong, older German adults are likely to use health apps if they are at the lower end of the age range and have a high level of technical readiness [13,14]. About 16.5% of questioned older adults already used at least one health app, which is comparable with the national survey results for populations in the United States and Hong Kong [13,14]. These results can be considered as reliable, as the reported response rate of our survey (11.52%, 576/5000) is consistent with methodical investigations of response rates among older adults

[18]. Regarding health app use, our findings extend current research results because besides the simple possession of health apps, we also investigated their use in terms of number of installed apps and frequency of use associated with type of app used among older adults in Germany.

Shen et al reported for the Hong Kong territory an association between history of chronic disease, as well as education and health app use [13]. In our study, neither association was confirmed. This might be related to the already high level of education within our sample. Comparing our participants' demographics with demographics of the general population in Germany reveals a higher level of education and higher representation of male individuals in our sample than would be expected for the population of adults older than 60 years (high level of education: 38.5% sample vs 16.5% general population; male gender: 51% sample vs 43% general population) [1,24]. Further efforts are needed to verify our results in a sample with individuals with lower levels of education, as they might have poor health status and therefore, would profit most from such apps [6,13]. A way to reach these groups are family and friends, as these are the most frequently used source of information about apps and therefore, would promote this topic best.

Krebs and Duncan reported that the group of exercise-related health apps is the most frequently used one among adults in the United States [15]. Our study showed consistent results as exercise-related health apps were also mentioned most frequently in our study. Further studies need to identify whether exercise-related apps are the initial point of contact with health-related apps or not. This would allow developers of health apps to get more insights on how to design initial contact and what prerequisites they could expect if developing health apps for older adults.

Identified barriers in our study were, in order of frequency, a lack of trust, data privacy concerns, and fear of misdiagnosis. This is a contrast to findings of Krebs and Duncan, where lack of interest, high cost, and lack of trust in apps collecting their data were mentioned most [15]. The costs of health apps were not a reason for older German adults to avoid the use of health apps, maybe because of the fact that most health care-related costs are paid by the health insurance companies, and this would also be expected for future health apps.

Limitations

This study has several limitations relating to its methodological design as well as the reported results. The postal study was not representative because of its self-selection. Although the selection of the original 5000 potential participants was adequate in terms of representative population characteristics, a bias is still possible as these decided by self-selection to take part in this study. A bias in recruitment will lead to differences for the groups in the use of health apps as well as in age and education. Due to the postal recruitment method, no inferences can be made about the frequency of use and sociodemographic distribution of the general user group of health apps, especially as this postal survey was conducted with older adults in Germany. Finally, it must be noted that we are unable to answer how strongly participants are engaged with their health apps in terms of time of use and adherence [25]. Although we know how frequently they consult their health apps, we did not investigate which specific tasks they use them for.

Conclusions

In an exploratory approach, we investigated the prevalence of health apps among older adults in Germany. A prevalence of health apps was identified, with exercise-related health apps being the most frequently used ones. We were able to determine barriers to and incentives for health app use and compared these with recent results from the United States and Hong Kong.

Acknowledgments

This publication is part of the research project "TECH4AGE," financed by the Federal Ministry of Education and Research (BMBF, under Grant No. 16SV7111) and promoted by VDI/VDE Innovation + Technik GmbH.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Introduction text of survey in German and English.

[PDF File (Adobe PDF File), 29KB - mhealth_v6i1e26_app1.pdf]

References

1. Statistisches Jahrbuch Deutschland 2016. Wiesbaden: Statistisches Bundesamt; 2016.
2. Rasche P, Schäfer K, Theis S, Bröhl C, Wille M, Schlick C, et al. Wandel von Technikakzeptanz und -Nutzung im Hohen Alter [Change of Technology Acceptance and Use in Old Age]. Z Gerontol Geriatr 2016 Sep;49 Suppl 1:1-152. [doi: 10.1007/s00391-016-1123-0] [Medline: 27554080]
3. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]

4. Klasnja P, Pratt W. Healthcare in the pocket: mapping the space of mobile-phone health interventions. J Biomed Inform 2012 Feb;45(1):184-198 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.017] [Medline: 21925288]
5. Wille M, Theis S, Rasche P, Bröhl C, Schlick C, Mertens A. Best practices for designing electronic healthcare devices and services for the elderly. i-com 2016;15(1). [doi: 10.1515/icom-2016-0009]
6. Eysenbach G. Consumer health informatics. Br Med J 2000 Jun 24;320(7251):1713-1716 [FREE Full text] [Medline: 10864552]
7. Kim H, Lee KH, Kim H, Kim JH. Using mobile phones in healthcare management for the elderly. Maturitas 2014 Dec;79(4):381-388. [doi: 10.1016/j.maturitas.2014.08.013] [Medline: 25270725]
8. Silva BM, Rodrigues JJ, de la Torre Diez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. J Biomed Inform 2015;56:265-272. [doi: 10.1016/j.jbi.2015.06.003]
9. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 09;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
10. Boyle L, Grainger R, Hall RM, Krebs JD. Use of and beliefs about mobile phone apps for diabetes self-management: surveys of people in a hospital diabetes clinic and diabetes health professionals in New Zealand. JMIR Mhealth Uhealth 2017 Jun 30;5(6):e85 [FREE Full text] [doi: 10.2196/mhealth.7263] [Medline: 28666975]
11. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
12. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The “Aachen fall prevention App” - a smartphone application app for the self-assessment of elderly patients at risk for ground level falls. Patient Saf Surg 2017;11(14). [Medline: 28503199]
13. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SS, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. JMIR Mhealth Uhealth 2017 Jun 05;5(6):e77 [FREE Full text] [doi: 10.2196/mhealth.7628] [Medline: 28583905]
14. Fox S, Duggan M. Pewinternet. 2012. Mobile health 2012 URL: http://www.pewinternet.org/2012/11/08/mobile-health-2012/ [accessed 2017-07-25] [WebCite Cache ID 6sDLewaQl]
15. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
16. Prensky M. Digital Natives, Digital Immigrants Part 1. On the Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
17. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
18. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]
19. Edelman LS, Yang R, Guymon M, Olson LM. Survey methods and response rates among rural community dwelling older adults. Nurs Res 2013;62(4):286-291. [doi: 10.1097/NNR.0b013e3182987b32] [Medline: 23817286]
20. Neyer FJ, Felber J, Gebhardt C. Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. Diagnostica 2012;58(2):87-99. [doi: 10.1026/0012-1924/a000067]
21. Sengpiel M, Dittberner D. The computer literacy scale (CLS) for older adults – development and validation. In: Herczeg M, Kindsmüller MC, editors. Mensch und Computer 2008: Viel Mehr Interaktion. München: Oldenbourg Verlag; 2008:7-16.
22. Röthlin F, Pelikan J, Ganahl K. lbihpr.lbg.ac.at.w8.netz-werk. Wien: Ludwig Boltzmann Gesellschaft GmbH; 2013. Die Gesundheitskompetenz von 15-jährigen Jugendlichen in Österreich URL: http://lbihpr.lbg.ac.at.w8.netz-werk.com/sites/files/lbihpr/attachments/hljugend_bericht.pdf [accessed 2018-01-18] [WebCite Cache ID 6wYyQW3zQ]
23. Bech M, Kristensen MB. Differential response rates in postal and Web-based surveys in older respondents. J Eur Surv Res Assoc 2009;3(1):1-6. [doi: 10.18148/srm/2009.v3i1.592]
24. destatis.de. Wiesbaden: Statistisches Bundesamt; 2015. Die Generation 65+ in Deutschland URL: https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2015/generation65/Pressebroschuere_generation65.pdf?__blob=publicationFile [accessed 2018-01-18] [WebCite Cache ID 6wYyhVbqN]
25. Mertens A, Brandl C, Miron-Shatz T, Schlick C, Neumann T, Kribben A, et al. A mobile application improves therapy-adherence rates in elderly patients undergoing rehabilitation: a crossover design study comparing documentation via iPad with paper-based control. Medicine (Baltimore) 2016 Sep;95(36):e4446 [FREE Full text] [doi: 10.1097/MD.0000000000004446] [Medline: 27603339]

Abbreviations

AOR: adjusted odds ratio
EU HLS-47: European Health Literacy Survey-47
QoL: quality of life

Edited by G Eysenbach; submitted 02.08.17; peer-reviewed by S Becker, P Krebs; comments to author 23.11.17; revised version received 05.12.17; accepted 18.12.17; published 23.01.18.

Please cite as:
Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, Mertens A
Prevalence of Health App Use Among Older Adults in Germany: National Survey
JMIR Mhealth Uhealth 2018;6(1):e26
URL: http://mhealth.jmir.org/2018/1/e26/
doi:10.2196/mhealth.8619
PMID:29362211

Original Paper

Health Information Technology Usability Evaluation Scale (Health-ITUES) for Usability Assessment of Mobile Health Technology: Validation Study

Rebecca Schnall[1], RN, MPH, PhD; Hwayoung Cho[1], RN, PhD; Jianfang Liu[1], BE, MAS, PhD

School of Nursing, Columbia University, New York, NY, United States

Corresponding Author:
Rebecca Schnall, RN, MPH, PhD
School of Nursing
Columbia University
560 West 168th Street
New York, NY, 10032
United States
Phone: 1 2123426886
Email: rb897@columbia.edu

Abstract

Background: Mobile technology has become a ubiquitous technology and can be particularly useful in the delivery of health interventions. This technology can allow us to deliver interventions to scale, cover broad geographic areas, and deliver technologies in highly tailored ways based on the preferences or characteristics of users. The broad use of mobile technologies supports the need for usability assessments of these tools. Although there have been a number of usability assessment instruments developed, none have been validated for use with mobile technologies.

Objective: The goal of this work was to validate the Health Information Technology Usability Evaluation Scale (Health-ITUES), a customizable usability assessment instrument in a sample of community-dwelling adults who were testing the use of a new mobile health (mHealth) technology.

Methods: A sample of 92 community-dwelling adults living with HIV used a new mobile app for symptom self-management and completed the Health-ITUES to assess the usability of the app. They also completed the Post-Study System Usability Questionnaire (PSSUQ), a widely used and well-validated usability assessment tool. Correlations between these scales and each of the subscales were assessed.

Results: The subscales of the Health-ITUES showed high internal consistency reliability (Cronbach alpha=.85-.92). Each of the Health-ITUES subscales and the overall scale was moderately to strongly correlated with the PSSUQ scales (r=.46-.70), demonstrating the criterion validity of the Health-ITUES.

Conclusions: The Health-ITUES has demonstrated reliability and validity for use in assessing the usability of mHealth technologies in community-dwelling adults living with a chronic illness.

(JMIR Mhealth Uhealth 2018;6(1):e4) doi:10.2196/mhealth.8851

KEYWORDS

mobile technology; usability; mobile health apps; psychometric evaluation

Introduction

Mobile technology has become a nearly ubiquitous technology in the United States, where almost two-thirds of the population owns a mobile phone, but also worldwide with more than half of the world population owning a mobile phone [1]. Mobile technology has the advantage of accessibility anywhere Internet access is available, relative affordability, and has been promoted specifically as a solution to reach stigmatized and disenfranchised populations [2,3]. Mobile technology can be particularly useful in the delivery of health interventions because this technology can allow us to deliver interventions to scale, cover broad geographic areas, and deliver technologies in highly tailored ways based on the preferences or characteristics of users [4,5].

Concurrent with the proliferation of mobile technology has been the vast increase in the number of mobile health (mHealth) apps

available to consumers. For example, in 2013, a report by the IMS Institute for Healthcare Informatics estimated more than 40,000 health-related iPhone apps were available for consumer use [6]. By 2015, this number had more than doubled to 90,088 apps available on the iOS. In total, IMS estimated that 165,000 health apps were available for use both among iOS and Android platforms [6]. The mHealth apps use mobile devices (eg, mobile phones or tablets) for the deployment of health interventions and vary widely in the types of technology used (eg, text message reminders, app-based alerts and activities, and Web-based educational modules) [7].

The dramatic growth of mHealth technology is not surprising given that mobile technologies are a novel method for the delivery of cost-effective, timely, and relevant health promotion and management information [8]. Mobile delivery specifically has a number of key advantages over traditional face-to-face delivery models of care, including consumer control, decreased time burden, reduction of monetary and time costs associated with travel to a provider, and the ability to monitor and assess the use of digital analytics [8]. Additionally, mHealth technology presents an opportunity for consumers to self-monitor their health status and allows health care providers to organically reach persons who may be disengaged from the health care system [9]. Finally, mHealth technology allows for the dissemination of information quickly and broadly [10].

Use of mobile technology as a platform for behavioral interventions adjuncts to health care delivery and monitoring of health status has widely proliferated in the past 5 years. Despite this growth, few mHealth apps or interventions have undergone systematic and rigorous usability evaluation prior to their dissemination [11,12]. Usability evaluations remain challenging because of the cost and time necessary for completion of these assessments [13].

Another barrier to usability assessments is that a dearth of validated instruments for assessing mHealth technology persists [14-16]. Implications of insufficient usability testing in mHealth include the assertion that attrition is common within randomized controlled trials using technology due to usability errors [11,12,17]. However, without a standardized measure of usability, it is very difficult to assess, report on, and compare usability, as well as link it to attrition from the trial. At the same time, a standardized usability instrument would improve methodological consistency, making it possible to begin comparing findings across mHealth app evaluations [18].

In our earlier work, we presented preliminary validation of the Health Information Technology Usability Evaluation Model (Health-ITUEM), a theoretical framework to guide usability evaluations for assessing mHealth technologies [19]. The model was developed in response to the current gaps in the extant usability literature [20]. The Health-ITUEM is an integrated model of multiple usability theories based on the concepts of usability from the Technology Acceptance Model [21] and the International Organization for Standardization (ISO) standard 9241-11 [22]. The Health-ITUEM has been widely used as a framework for understanding the use of mobile technology studies since its validation [23,24], but there remains a lack of validated instruments for usability assessments of mHealth technology.

A need for a validated usability instrument for mHealth technologies persists. In response to this need, we sought to validate the use of the Health Information Technology Usability Evaluation Scale (Health-ITUES) in a sample of users of a mHealth app developed for community-dwelling adults living with HIV. The Health-ITUES, derived from the Health-ITUEM, is a validated instrument that explicitly considers task by addressing various levels of expectation of support for the task by the health information technology (IT). The Health-ITUES also has the added benefit of being customizable, which can address the study needs and concepts measured without item addition, deletion, or modification. This is an important benefit of this instrument because it allows for harmonization of findings across studies. The factorial validity and internal consistency of the Health-ITUES was demonstrated through an exploratory factor analysis [25]. The development of the Health-ITUES was further advanced through a confirmatory factor analysis and structural equation modeling to demonstrate its construct validity and predictive validity [26]. Early development and valuation of the Health-ITUES was conducted using a Web-based communication system that supported nurse staffing and scheduling. The Health-ITUES is not widely used and it has not been validated in a sample of patients or with the use of mobile technology. These two distinctions make the findings from this work highly relevant, timely, and provide an important contribution to the literature.

Setting

These data were collected as part of a larger study for developing and testing a mobile app (mVIP) for symptom self-management in persons living with HIV. The goal of the parent study was to translate a paper-based manual that had been effective in ameliorating symptoms in persons living with HIV onto a mobile platform to promote dissemination of patient-centered outcomes research. As part of the parent study, we developed a beta version of the app and conducted end-user usability testing with 20 participants in a laboratory setting. Following refinement of the app based on our findings during our usability assessment, we tested the mobile app in a 12-week feasibility trial with 80 end users. Participants were randomized to an intervention or control group, but both groups received a version of the mobile app. The intervention group received an enhanced app, which included self-care strategies for 13 symptoms commonly experienced by persons living with HIV.

Methods

Study participants were recruited from June 2016 to February 2017 through in-person and email recruitment. Recruitment sites included HIV clinics and community-based organizations. Eligibility criteria included: (1) diagnosed with HIV, (2) older than 18 years of age, (3) able to communicate in English, (4) experienced at least two of 13 HIV-related symptoms in the past week, (5) met the cognitive state minimum score (ie, 24 of 30) measured by the Mini-Mental State Examination (MMSE) [27], and (6) own a mobile phone or tablet. All participants completed written informed consent prior to the start of study

activities. Following enrollment, study participants were given access to the mVIP app. A total of 20 participants enrolled as part of usability testing of the app and 80 participants enrolled in the 12-week feasibility study of the app. Data reported in this paper were collected at the baseline visit for each participant after he or she had used the app for the first time.

Participants in both groups were given access to the app on a mobile phone. Following log-in and use of the mobile app, participants completed a number of surveys, which are detailed subsequently. The surveys collected demographic information, the Health-ITUES [26], and the Post-Study System Usability Questionnaire (PSSUQ) [28]. Participants were given US $40 as a token of appreciation for their time for participating in the usability testing. Participants in the 12-week feasibility trial were given US $30 at baseline and US $40 at follow-up as a token of appreciation for their time.

Instruments

Health Information Technology Usability Evaluation Scale

The Health-ITUES is a customizable questionnaire with a four-factor structure. The Health-ITUES explicitly considers a task by addressing various levels of expectation of support for the task by the health IT system, and it has been validated through exploratory factor analysis and confirmatory factor analysis in a sample of nurses who used a Web-based nurse-scheduling tool. The Health-ITUES consists of 20 items rated on a five-point Likert scale from strongly disagree (1) to strongly agree (5). A higher scale value indicates higher perceived usability of the technology. Items in each of these scales and how they were customized for this study are illustrated in Figure 1. The 20-item scale is comprised of four subscales: (1) quality of work life, (2) perceived usefulness, (3) perceived ease of use, and (4) user control. User control and perceived ease of use capture user-system interaction, whereas perceived usefulness evaluates task accomplishment through system use and quality of work life represents higher expectations of system impact. In the earlier studies, quality of work life referred to the system impact on work life; in our study with community-dwelling persons living with HIV, quality of work life represents the system impact on daily life. As a result, we renamed this factor structure for purposes of this validation as “impact.” The overall Health-ITUES score was the mean of all the items with each item weighted equally.

Sociodemographic Questionnaire

A self-reported sociodemographic questionnaire was developed to collect information on the participants’ age, gender, race, ethnicity, sexual orientation, education, annual income, relationship status, and a previous acquired immune deficiency syndrome (AIDS) diagnosis. Participants were also asked about their technology use with questions specifically asking about most frequently used type of mobile device, frequency of use of this device, and whether they use apps on their mobile phone.

Post-Study System Usability Questionnaire

The PSSUQ is an instrument for assessing user satisfaction with system usability and was developed as a usability assessment tool that was specifically for use in the context of scenario-based usability testing, although additional research has indicated that this may be useful for field evaluation as well [29,30]. Factor analysis of PSSUQ support a three-factor structure: system usefulness, information quality, and interface quality. For this study, we used the PSSUQ version 3, which is comprised of 16 items [28]. The items are rated on a seven-point scale, anchored at the end points with the terms strongly agree (1), strongly disagree (7), and a “not applicable” (N/A) point outside the scale [28,31]. The overall PSSUQ and subscale scores were reversely coded in this study so that higher scores indicated better user satisfaction. As a follow-up to the original PSSUQ, the developers of the instrument collected data from 5 years of usability studies and found similar psychometric properties between the original and the follow-up PSSUQ data despite the passage of time and differences in the types of systems studied, providing evidence of significant generalizability of the instrument for measuring participant satisfaction with the usability of tested systems [28].

Procedures

Descriptive statistics were used to calculate demographic characteristics of the study sample. We used the Kruskal-Wallis test to determine if there was a relationship between any of the demographic data and each of the Health-ITUES constructs: impact, perceived usefulness, perceived ease of use, and user control.

Psychometric test theory involves the development and evaluation of clusters of questions called scales, which are used to gather information about the quality of psychological measures [32]. The Health-ITUES has been previously tested using exploratory factor analysis, confirmatory factor analysis, and structural equation modeling, as described previously, but it had not been validated in patients or with mobile technology. Therefore, there is a need for a further psychometric evaluation of this instrument. We evaluated the following properties in our study sample: variability, internal consistency reliability, construct validity, and criterion validity. Each property and the appropriate analysis are described subsequently. Outliers were checked and removed in the reliability and validity analysis.

Variability refers to the extent to which the full range of scale scores and item responses are reported in the data. Optimal variability is denoted by a full range of responses of the scale. Scales that are skewed, either negatively or positively, tend to be less responsive to the effect of usability errors. To ensure limits in the variability, the frequency of missing data needs to be limited and randomly distributed across participant responses. To address skewness, we used a nonparametric method (Kruskal-Wallis) to test the known-group validity. For the other analyses, such as Cronbach alpha and Pearson correlation, normality is not an assumption, so no log transformation was performed in this analysis.

Figure 1. Untitled.

Health-ITUES

	STRONGLY AGREE 5 4 3 2 1 STRONGLY DISAGREE
Impact	
1	I think mVIP would be a positive addition for persons living with HIV.
2	I think mVIP would improve the Quality of Life of persons living with HIV.
3	mVIP is an important part of meeting my information needs related to symptom self-management.
Perceived Usefulness	
4	Using mVIP makes it easier to self-manage my HIV-related symptoms.
5	Using mVIP enables me to self-manage my HIV-related symptoms more quickly.
6	Using mVIP makes it more likely that I can self-manage my HIV-related symptoms.
7	Using mVIP is useful for self-management of HIV-related symptoms.
8	I think mVIP presents a more equitable process for self-management of HIV-related symptoms.
9	I am satisfied with mVIP for self-management of HIV-related symptoms.
10	I self-manage my HIV-related symptoms in a timely manner because of mVIP.
11	Using mVIP increases my ability to self-manage my HIV-related symptoms.
12	I am able to self-manage my HIV-related symptoms whenever I use mVIP.
Perceived Ease of Use	
13	I am comfortable with my ability to use mVIP.
14	Learning to operate mVIP is easy for me.
15	It is easy for me to become skillful at using mVIP.
16	I find mVIP easy to use.
17	I can always remember how to log on to and use mVIP.
User Control	
18	mVIP gives error messages that clearly tell me how to fix problems.
19	Whenever I make a mistake using mVIP, I recover easily and quickly.
20	The information (such as on-line help, on-screen messages and other documentation) provided with mVIP is clear.

Internal consistency reliability is a measure of the how well the instrument measures different constructs and delivers reliable scores. Internal consistency reliability measures whether several items that propose to measure the same general construct produce similar scores, indicating the homogeneity of a scale or subscale. The Cronbach alpha coefficient provides an estimated score of the internal consistency reliability based on all possible correlations between items collected at any time point [33]. Cronbach alpha scores range from .0 (no reliability) to 1.0 (perfect reliability) with the desired range of scores between .70 and .95 [34]. Cronbach alphas greater than .95 demonstrate highly correlated items, which is not desired.

Validity refers to how well the scale measures the attribute it claims to measure [35,36]. There are several components to validity; in this study, we focused on the construct and criterion validity. We measured three subtypes of construct validity in our study: convergent validity, discriminant validity, and known-group validity. Convergent validity refers to the degree to which theoretically correlated measures are in fact correlated, whereas discriminant validity is used to evaluate the differences between uncorrelated and correlated subscales [37]. One way to examine the convergent and discriminant validity is to assess the correlations among scale scores within the instrument based on known relationships. For example, scales measuring perceived ease of use are expected to correlate moderately with one another, whereas scales measuring impact are expected to have weaker correlations with user control because they measure different constructs. We used a multitrait-multimethod matrix [38] with interscale correlations to assess the convergent and discriminant validity. We also assessed the known-groups validity, which tests for anticipated differences on specific scale scores between groups that are known to be different [39]. In

the case of our study population, we evaluated the ability of this instrument to distinguish between participants who were randomized to the intervention group compared to those who were randomized to the control group. We hypothesized that the intervention group participants, who received access to an enhanced app, would report higher usability scores driven by an increased subscore of usefulness compared to the control group participants.

Criterion validity is the extent to which a measure is correlated with a validated outcome measure [40]. To assess criterion validity, we used the correlation between the PSSUQ subscale and overall scores and the Health-ITUES subscale and overall scores[41].

Results

A total of 92 persons completed the Health-ITUES survey. Twenty participants completed the survey as part of the end-user usability testing of the beta version of the app and 72 persons completed the survey on enrollment into the 12-week trial. As noted in Table 1, the majority of participants were African American/black and had an annual median household income of less than US $20,000. Nearly all our participants used their mobile phones several times per day. The mean age of the participants was 50.0 (SD 10.5) years with a range of 23.0 to 72.0 years. There was no significant correlation between demographic characteristics, such as income and education, and each of the Health-ITUES subscales.

The range, mean, median, and standard deviation for each of the subscales' scores at baseline are reported in Table 2. Completion rates were identical for all subscales.

Variability was evaluated for each of the subscales. The full range of responses was not observed for any of the Health-ITUES subscales. Slightly more than a half (59%) scored the maximum score for the quality of life scale, a third (36%) scored the maximum score for the perceived usefulness scale, slightly more than a half (60%) scored the maximum score for the perceived ease of use scale, and more than a third (42%) scored the maximum score for the user control scale. No users scored the lowest possible score. Overall, the users rated the usability of mVIP app as being high. The full range of usability scores was not demonstrated because no users rated the app as the lowest possible score.

By examining the boxplots and the raw data, extremely low scores were treated as outliers and were removed from the reliability and validity analysis. There is evidence that all the subscales were negatively skewed, indicating more favorable usability scores.

Internal Consistency Reliability and Construct Validity

Internal consistency reliability, as measured by the Cronbach alpha coefficient, is reported for each of the multi-item scales in Table 3. All scales displayed very good Cronbach alpha values (>.7) with the scores ranging from .85 to .92. All the Cronbach alpha values were less than .95, which indicates no redundancy in items [32]. Interscale correlations were all less than the corresponding Cronbach alpha values and ranged from .56 to .82, which indicates moderate to strong correlations [42]. Impact was more correlated with perceived usefulness than with the perceived ease of use and user control (r=.82, r=.63, and r=.56, respectively), which is evidence of convergent validity and discriminant validity.

Known-groups validity, another subtype of construct validity, was evaluated by measuring differences in mean scale scores at baseline between the control and intervention group (Table 4). As noted previously, the subscale scores were skewed and thus the assumption of normality might be violated for independent samples t test, therefore we used a nonparametric method, the Kruskal-Wallis test, to detect differences between the intervention and control group scores. Using the Kruskal-Wallis test, we found a statistically significant difference between the intervention and control group on all subscales expect impact.

Criterion Validity

Criterion validity was measured through the assessment of the concurrent validity by comparing the measure in question, the Health-ITUES, with an outcome assessed at the same time, the PSSUQ. Table 5 presents information on criterion validity, measured as the correlation between each of the subscales of the Health-ITUES with each of the subscales of a validated usability measurement tool, the PSSUQ. Each of the Health-ITUES subscales and the overall scale were moderately to strongly correlated with the PSSUQ scales. These correlations were significant at the P<.001 level. Correlations ranged from .46 to .70. None of the correlations were greater than .80 or were very strong and thus there was no redundancy in subscales [42]. The least correlated subscales of the measures were impact and information quality and the most highly correlated items were the overall scale scores.

Table 1. Participant characteristics (N=92).

Characteristics	n (%)
Gender	
Male	50 (54)
Female	40 (43)
Transgender (FTM)	1 (1)
Other	1 (1)
Race	
African American/black	67 (74)
White	8 (9)
Other	16 (18)
Ethnicity	
Hispanic	19 (21)
Non-Hispanic	73 (79)
Sexual orientation	
Homosexual/gay/lesbian	24 (33)
Heterosexual/straight	44 (61)
Bisexual	3 (4)
Other	1 (1)
Education	
Elementary	1 (1)
Some high school	16 (17)
High school diploma or equivalent	27 (29)
Some college	28 (30)
Associate/technical degree	5 (5)
Bachelor/college degree	15 (16)
Annual income (US$)	
Less than $10,000	41 (45)
$10,000-$19,999	24 (26)
$20,000-$39,999	14 (15)
$40,000-$59,999	1 (1)
$60,000-$79,999	1 (1)
$80,000-$99,999	1 (1)
Don’t know	5 (5)
Prefer not to answer	5 (5)
Relationship status	
Married or in a steady relationship	22 (31)
Single, separated, divorced, or widowed	48 (67)
Other	2 (3)
Most frequently used mobile devices	
Android phone	62 (67)
iPhone	23 (25)
Tablet	5 (5)
Other	2 (2)

Characteristics	n (%)
Frequency of using the mobile device	
Several times every day	82 (89)
Once a day	5 (5)
Several times per week	4 (4)
Several times per month	1 (1)
Download/use apps on the mobile device	
Yes	80 (88)
No	12 (13)

Table 2. Descriptive statistics: scale scores at enrollment for the Health-ITUES subscales (N=92).

Scale	Mean (SD)	Median (range)	Floor, %	Ceiling, %
Impact	4.5 (0.7)	5.0 (1.3-5.0)	0	59
Perceived usefulness	4.3 (0.8)	4.5 (1.7-5.0)	0	36
Perceived ease of use	4.6 (0.8)	5.0 (1.2-5.0)	0	60
User control	4.2 (0.9)	4.5 (1.7-5.0)	0	42
Overall Health-ITUES score	4.4 (0.7)	4.6 (1.6-5.0)	0	33

Table 3. Internal scale consistency scores and interscale correlations for Health-ITUES subscales (N=83).

Scale	Impact		Perceived usefulness		Perceived ease of use		User control	
	Cronbach alpha	*r*	Cronbach alpha	*r*	Cronbach alpha	*r*	Cronbach alpha	*r*
Impact	.85							
Perceived usefulness		.82	.92					
Perceived ease of use		.63		.69	.92			
User control		.56		.68		.61	.86	

Table 4. Mean scale scores at baseline by intervention versus control groups for Health-ITUES subscales (N=83).

Scale	Intervention (n=49) Mean (SD)	Control (n=34) Mean (SD)	*P* [a]
Impact	4.70 (0.52)	4.53 (0.61)	.13
Perceived usefulness	4.58 (0.59)	4.28 (0.60)	.01
Perceived ease of use	4.77 (0.48)	4.52 (0.64)	.03
User control	4.45 (0.73)	4.11 (0.82)	.048
Overall Health-ITUES score	4.62 (0.51)	4.35 (0.56)	.01

[a]From Kruskal-Wallis test.

Table 5. Correlations between Health-ITUES and PSSUQ subscales (N=83).

PSSUQ	Health-ITUES, *r*[a]				
	Impact	Perceived usefulness	Perceived ease of use	User control	Overall
System usefulness	.63	.53	.57	.49	.63
Information quality	.46	.56	.52	.67	.65
Interface quality	.56	.49	.50	.55	.61
Overall	.60	.60	.59	.64	.70

[a]All correlations significant at the *P*<.001 level.

Discussion

This paper presents the validation of an instrument for measuring usability of mHealth technology. The validation study reported in this paper was conducted to explore the psychometric properties of the Health-ITUES, a customizable usability evaluation instrument that includes subscales of impact, perceived usefulness, ease of use, and user control in a sample of community-dwelling adults living with HIV who were users of a mobile app for symptom self-management. The Health-ITUES varies from most traditional measurement scales in that it is designed to support customization at the item level to match the specific task/expectation and health IT system while retaining comparability at the construct level. The results of this study support the construct and criterion validity and reliability of the Health-ITUES for usability assessments of mobile technologies.

In this study, the reliabilities of the four Health-ITUES subscales and overall score were examined in terms of internal consistency. The internal consistency reliability of each of the subscales were high but less than the Cronbach alpha=.95 threshold suggesting that there is no indication of redundancy in scale items. Moreover, a comparison of the results from the Health-ITUES to the PSSUQ had moderate to strong correlations (r=.46-.70). Information quality, a subscale on the PSSUQ, had the lowest correlation (r=.46, P<.001), which is to be expected because the constructs are substantively different. The overall scale scores for the PSSUQ and the Health-ITUES had the highest correlation (r=.70, P<.001). This would be expected because the two usability scales have some overlap in their constructs. At the same time, the overall correlations were only moderately strong [42], suggesting that these two instruments are not redundant and demonstrating good criterion validity.

A strength of this validation is our study sample, which further supports the utility of this instrument for usability assessments for consumer health informatics. Our study sample is comprised of nearly all racial and ethnic minority persons and those persons of the lowest socioeconomic groups in the United States. Nonetheless, our study participants were owners and frequent users of mobile phone devices, demonstrating the widespread use of mobile technology and the relevance of mHealth technology across end users. In contrast to the earlier development and validations of the Health-ITUES in nurses using a Web-based system [25,26], this paper presents a validation of this instrument for mHealth technology and with patients [43].

Construct validity was assessed by measuring differences in mean scale scores at baseline between control group and intervention group participants. Statistically significant differences between groups were found in the expected direction for all subscales except impact. These trends were expected because at baseline we would not expect to find a significant difference in impact as a result of the study intervention. In contrast, a statistically significant difference in perceived usefulness, ease of use, and user control was demonstrated, with intervention group participants finding the mobile app to be more usable than control group participants did.

The construct validity of the measure was also supported by the degree to which interscale correlations corresponded to what was expected. The higher correlations among the impact and perceived usefulness subscales than between the other subscale items supports the convergent and discriminant validity of this instrument because participants who perceive the technology to be useful are more likely to show an improvement in their impact as a response.

There are a number of limitations of this validation study. First, most of our study sample were mobile phone users and so we were unable to validate the use of this instrument as a usability assessment tool in non-mobile phone users. This limitation is mitigated as the number of mobile phone users continues to grow across geographic regions and socioeconomic groups. A second limitation is that these findings may not be generalizable beyond our study population. Further research is most definitely needed to further validate this instrument in other study populations. Given the dearth of current instruments for usability assessments, we believe that the findings from this study provide an important contribution to the literature and an opportunity for future study of this instrument with additional study populations.

This study provides preliminary evidence to support the validity and reliability of the Health-ITUES for usability assessments of mobile technology. In light of our findings, the authors recommend use of the Health-ITUES as a measurement tool for assessing the usability of mobile technologies. Validation of the Health-ITUES is an important step in ensuring that usability is attended to and maintained in mHealth technology interventions. This work is particularly important given the proliferation of mobile technologies and the push for consumers to take control of their own health and become greater users of technology for the acquisition and delivery of health information. Further, these findings will hopefully stimulate additional research and practice to ensure the usability of mobile technologies and particularly for those tools that are developed for consumers' use.

Acknowledgments

Research reported in this publication was support by the US Department of Health and Human Services, Agency for Healthcare Research and Quality under award number R21HS023963. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality.

Conflicts of Interest

None declared.

References

1. We Are Social. Digital in 2017: global overview 2017 URL: https://wearesocial.com/special-reports/digital-in-2017-global-overview [accessed 2017-12-16] [WebCite Cache ID 6vlisUSt1]
2. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
3. Devi BR, Syed-Abdul S, Kumar A, Iqbal U, Nguyen P, Li YJ, et al. mHealth: an updated systematic review with a focus on HIV/AIDS and tuberculosis long term management using mobile phones. Comput Methods Programs Biomed 2015 Nov;122(2):257-265. [doi: 10.1016/j.cmpb.2015.08.003] [Medline: 26304621]
4. Eysenbach G, Wyatt J. Using the Internet for surveys and health research. J Med Internet Res 2002;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
5. Wyatt JC. When to use web-based surveys. J Am Med Inform Assoc 2000;7(4):426-429 [FREE Full text] [Medline: 10887170]
6. IMS Institute for Healthcare Informatics. Availability and profile of consumer mHealth apps 2015 URL: https://www.iqvia.com/ [accessed 2017-12-16] [WebCite Cache ID 6vlizsv8T]
7. Iribarren S, Schnall R. Health Informatics: An Interprofessional Approach. St Louis, MO: Elsevier; 2017.
8. Schnall R, Cho H, Webel A. Predictors of willingness to use a smartphone for research in underserved persons living with HIV. Int J Med Inform 2017 Dec;99:53-59. [doi: 10.1016/j.ijmedinf.2017.01.002] [Medline: 28118922]
9. Schnall R, Bakken S, Rojas M, Travers J, Carballo-Dieguez A. mHealth technology as a persuasive tool for treatment, care and management of persons living with HIV. AIDS Behav 2015 Jun;19 Suppl 2:81-89 [FREE Full text] [doi: 10.1007/s10461-014-0984-8] [Medline: 25572830]
10. mHealth: New horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2017-12-16] [WebCite Cache ID 6vljg8S6M]
11. Furlow B. Clinical Advisor. 2012 Nov 02. mHealth apps may make chronic disease management easier 2012 URL: http://www.clinicaladvisor.com/features/mhealth-apps-may-make-chronic-disease-management-easier/article/266782/ [accessed 2017-12-16] [WebCite Cache ID 6vlkhSyHX]
12. Zapata B, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
13. Holzinger A. Usability engineering methods for software developers. Commun ACM 2005 Jan 01;48(1):71-74. [doi: 10.1145/1039539.1039541]
14. Schnall R, Rojas M, Travers J, Brown W, Bakken S. Use of design science for informing the development of a mobile app for persons living with HIV. AMIA Annu Symp Proc 2014;2014:1037-1045 [FREE Full text] [Medline: 25954413]
15. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
16. Sheehan B, Lee Y, Rodriguez M, Tiase V, Schnall R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. Appl Clin Inform 2012;3(4):356-366 [FREE Full text] [doi: 10.4338/ACI-2012-06-RA-0021] [Medline: 23227134]
17. Maguire M. Methods to support human-centred design. Int J Hum-Comput St 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
18. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
19. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
20. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
21. Davis F, Bagozzi R, Warshaw P. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
22. International Organization for Standardization. Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11: Guidance on usability1998 URL: https://www.iso.org/standard/16883.html [accessed 2017-12-16] [WebCite Cache ID 6vllBMK8Z]
23. Househ MS, Shubair MM, Yunus F, Jamal A, Aldossari B. The Use of an Adapted Health IT Usability Evaluation Model (Health-ITUEM) for evaluating consumer reported Ratings of Diabetes mHealth Applications: Implications for Diabetes

Care and Management. Acta Inform Med 2015 Oct;23(5):290-295 [FREE Full text] [doi: 10.5455/aim.2015.23.290-295] [Medline: 26635437]
24. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth Quality: a process to seal the qualified mobile health apps. Stud Health Technol Inform 2016;228:205-209. [Medline: 27577372]
25. Yen P, Wantland D, Bakken S. Development of a customizable health IT usability evaluation scale. AMIA Annu Symp Proc 2010 Nov 13;2010:917-921 [FREE Full text] [Medline: 21347112]
26. Yen P, Sousa K, Bakken S. Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. J Am Med Inform Assoc 2014 Oct;21(e2):e241-e248 [FREE Full text] [doi: 10.1136/amiajnl-2013-001811] [Medline: 24567081]
27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
28. Lewis J. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum-Comput Int 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
29. Lewis J. Psychometric evaluation of the post-study system usability questionnaire: the PSSUQ. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1992 Presented at: Human Factors and Ergonomics Society Annual Meeting; Oct 12-16, 1992; Atlanta, GA p. 1259-1260. [doi: 10.1177/154193129203601617]
30. Lewis J, Henry S, Mack R. Integrated office software benchmarks: a case study. In: Proceedings of the third IFIP Conference on Human-Computer Interaction–INTERACT. 1990 Presented at: IFIP Conference on Human-Computer Interaction–INTERACT; Aug 27-30, 1990; Cambridge, UK p. 43.
31. Sauro J, Lewis J. Quantifying the User Experience: Practical Statistics for User Research. Burlington, MA: Morgan Kaufmann; 2016.
32. Nunnally J, Bernstein IH. Psychometric Theory. New York: McGraw-Hill; 1994.
33. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16. [doi: 10.1007/BF02310555]
34. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
35. National Council on Measurement in Education. URL: http://www.ncme.org/ncme/NCME/Resource_Center/Glossary/NCME/Resource_Center/Glossary1.aspx?hkey=4bb87415-44dc-4088-9ed9-e8515326a061 [accessed 2017-12-17] [WebCite Cache ID 6vlmrYdjY]
36. Brains C. Empirical Political Analysis. Abingdon, UK: Routledge; 2011:105.
37. Cronbach L, Meehl P. Construct validity in psychological tests. Psychol Bull 1955;52(4):281-302.
38. Campbell DT, Fiske D. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol Bull 1959;56(2):81-105. [doi: 10.1037/h0046016]
39. National Survey of Student Engagement. URL: http://nsse.indiana.edu/html/validity.cfm [accessed 2017-12-16] [WebCite Cache ID 6vlk3ahfE]
40. Standards for Educational and Psychological Testing. Washington, DC: American Psychological Association; 1985.
41. Justice A, Holmes W, Gifford A, Rabeneck L, Zackin R, Sinclair G. Development and validation of a self-completed HIV symptom index. J Clin Epidemiol 2001;54(12, Supplement 1):S77-S90. [Medline: 11750213]
42. Evans J. Straightforward Statistics for the Behavioral Sciences. Pacific Grove, CA: Brooks/Cole Pub Co; 1996.
43. Deeks S, Lewin S, Havlir D. The end of AIDS: HIV infection as a chronic disease. Lancet 2013 Nov 02;382(9903):1525-1533 [FREE Full text] [doi: 10.1016/S0140-6736(13)61809-7] [Medline: 24152939]

Abbreviations

AIDS: acquired immune deficiency syndrome
Health-ITUEM: Health Information Technology Usability Evaluation Model
Health-ITUES: Health Information Technology Usability Evaluation Scale
IT: information technology
mHealth: mobile health
PSSUQ: Post-Study System Usability Questionnaire

Edited by G Eysenbach; submitted 28.08.17; peer-reviewed by P Santana-Mancilla, S Baptista, M Kim; comments to author 29.09.17; revised version received 28.10.17; accepted 16.11.17; published 05.01.18.

Please cite as:
Schnall R, Cho H, Liu J
Health Information Technology Usability Evaluation Scale (Health-ITUES) for Usability Assessment of Mobile Health Technology: Validation Study
JMIR Mhealth Uhealth 2018;6(1):e4
URL: http://mhealth.jmir.org/2018/1/e4/
doi:10.2196/mhealth.8851
PMID:29305343

Original Paper

Concussion Assessment With Smartglasses: Validation Study of Balance Measurement Toward a Lightweight, Multimodal, Field-Ready Platform

Joseph P Salisbury[1,2], PhD; Neha U Keshav[2], PhD; Anthony D Sossong[1,2,3,4], MD; Ned T Sahin[1,2,5], PhD

[1]Neural Sensing and Biometrics Division, TIAX LLC, Lexington, MA, United States

[2]Empowerment Lab, Brain Power, LLC, Cambridge, MA, United States

[3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[4]Harvard Medical School, Boston, MA, United States

[5]Department of Psychology, Harvard University, Cambridge, MA, United States

Corresponding Author:
Ned T Sahin, PhD
Empowerment Lab
Brain Power, LLC
1 Broadway 14th Fl
Cambridge, MA, 02142
United States
Phone: 1 617 758 4100
Email: sahin@post.harvard.edu

Abstract

Background: Lightweight and portable devices that objectively measure concussion-related impairments could improve injury detection and critical decision-making in contact sports and the military, where brain injuries commonly occur but remain underreported. Current standard assessments often rely heavily on subjective methods such as symptom self-reporting. Head-mounted wearables, such as smartglasses, provide an emerging platform for consideration that could deliver the range of assessments necessary to develop a rapid and objective screen for brain injury. Standing balance assessment, one parameter that may inform a concussion diagnosis, could theoretically be performed quantitatively using current off-the-shelf smartglasses with an internal accelerometer. However, the validity of balance measurement using smartglasses has not been investigated.

Objective: This study aimed to perform preliminary validation of a smartglasses-based balance accelerometer measure (BAM) compared with the well-described and characterized waist-based BAM.

Methods: Forty-two healthy individuals (26 male, 16 female; mean age 23.8 [SD 5.2] years) participated in the study. Following the BAM protocol, each subject performed 2 trials of 6 balance stances while accelerometer and gyroscope data were recorded from smartglasses (Glass Explorer Edition). Test-retest reliability and correlation were determined relative to waist-based BAM as used in the National Institutes of Health's Standing Balance Toolbox.

Results: Balance measurements obtained using a head-mounted wearable were highly correlated with those obtained through a waist-mounted accelerometer (Spearman rho, ρ=.85). Test-retest reliability was high (intraclass correlation coefficient, $ICC_{2,1}$=0.85, 95% CI 0.81-0.88) and in good agreement with waist balance measurements ($ICC_{2,1}$=0.84, 95% CI 0.80-0.88). Considering the normalized path length magnitude across all 3 axes improved interdevice correlation (ρ=.90) while maintaining test-retest reliability ($ICC_{2,1}$=0.87, 95% CI 0.83-0.90). All subjects successfully completed the study, demonstrating the feasibility of using a head-mounted wearable to assess balance in a healthy population.

Conclusions: Balance measurements derived from the smartglasses-based accelerometer were consistent with those obtained using a waist-mounted accelerometer. Additional research is necessary to determine to what extent smartglasses-based accelerometry measures can detect balance dysfunction associated with concussion. However, given the potential for smartglasses to perform additional concussion-related assessments in an integrated, wearable platform, continued development and validation of a smartglasses-based balance assessment is warranted. This approach could lead to a wearable platform for real-time assessment of concussion-related impairments that could be further augmented with telemedicine capabilities to integrate professional clinical

guidance. Smartglasses may be superior to fully immersive virtual reality headsets for this application, given their lighter weight and reduced likelihood of potential safety concerns.

(JMIR Mhealth Uhealth 2018;6(1):e15) doi:10.2196/mhealth.8478

KEYWORDS

postural balance; wearable technology; accelerometry; mild traumatic brain injury

Introduction

Background

Mild traumatic brain injury (mTBI), also known as concussion, is a common injury in both sports, with an estimated annual incidence of 1.6-3.8 million in the United States alone [1], and modern war, with 297,478 diagnoses in US service members between 2000 and 2016 [2]. Prompt identification of a concussed individual and removal from activity is the most effective method to facilitate rapid recovery immediately following injury [3-6]. Unrecognized and untreated mTBI can put athletes and service members at greater risk for more substantial TBI, as well as chronic encephalopathy, later [7-10]. Unfortunately, failure to detect concussions in a timely fashion is common in both the sporting arena [11,12] and military [13], as the immediate symptoms can be subtle and difficult to detect.

Concussion is considered one of the most complex injuries in sports medicine to diagnose, assess, and manage [14]. Accurate diagnosis and recovery monitoring of concussion is further complicated as recommended assessments, including the Standardized Concussion Assessment Tool (SCAT) and Military Acute Concussion Evaluation (MACE), rely heavily on patient symptom self-reporting [15]. Concussions can escape detection in committed athletes who are motivated to remain in the game [16], which further highlights the need for unbiased and objective sideline assessments [17]. In a military setting, service members who experience concussions are frequently under severe levels of physiological and emotional stress and may be unable to recognize or recall symptoms [13]. Injuries are frequently embedded in longer, continuous missions, where removing oneself from active combat to report a mild injury often does not occur [13]. Furthermore, concussion assessments commonly used in these settings, including MACE, lack diagnostic utility as early as 12 hours after injury [18]. Thus, improved, more objective methods for detection and recovery monitoring following concussions are a priority for both athletic organizations [19-21] and US Department of Defense health care providers [22-24].

Approach

Concussion diagnosis and recovery monitoring requires a multifaceted and multimodal approach [25]. Concussion results in a range of clinical signs and symptoms, including impaired movement, balance, oculomotor function, attention, memory, and emotional functioning [26]. Unbiased and objective assessments of reaction time, balance, oculomotor function, and heart rate variability using an automated, digitized platform could substantially enhance the field recognition of concussion [25,27]. Although any single measure may not be precise enough to confidently diagnose concussion, a standardized combination of these measures could produce a sufficient concussion diagnostic metric. A lightweight and portable tool combining appropriate measures in a rapid assessment battery would be useful in both contact sports and the battlefield, where fast-paced and disorganized environments often obscure incidents of injury [25,28].

Considering the variety of assessments necessary, we examined whether smartglasses, an emerging computing platform, could be leveraged to provide a lightweight, portable, and wearable solution for measuring concussion-related impairments. Smartglasses, such as Glass (Google/X, Mountain View, CA), typically have a built-in 9-axis inertial measurement unit (IMU) that includes a 3-axis accelerometer, along with a gyroscope and magnetometer. Accelerometer-based balance assessments have garnered increased attention because of the widespread availability of accelerometers as a component of consumer smartphones [29,30]. The balance accelerometer measure (BAM) was developed as part of the National Institutes of Health's (NIH) Standing Balance Toolbox to provide a low-cost assessment [31], which can be administered through the use of an iOS app. Likewise, the Sway balance app [32-36] for iOS was designed to provide an easily accessible method for quantitative balance assessment and has obtained FDA (Food and Drug Administration) clearance.

Smartglasses could enable self-administered balance assessments, as well as rehabilitative feedback, by providing real-time audio/visual instruction to the user while monitoring balance via the IMU. Although balance would only be one component of a concussion diagnostic metric, smartglasses could also deliver other relevant assessments, including vestibulo-ocular and cognitive assessments. Smartglasses could also serve as a processing hub for integration with other wearable sensors, including wearable electrophysiological devices. Finally, smartglasses could enable remote/telemedicine concussion diagnosis. A medical professional could receive data from the wearable sensor platform while communicating in real time with the injured or some untrained personnel to determine the need for further care [37].

Goal of This Study

In this report, we sought to determine to what extent smartglasses-based balance measurement corresponded with a consumer smartphone attached at the waist, as in the NIH Toolbox Standing Balance Test. The objective of this study was to demonstrate the feasibility of obtaining quantitative balance measurements with smartglasses. These results could motivate future research in how smartglasses may be used to measure balance dysfunction and other concussion-related impairments. Although there exist multiple static balance protocols, the NIH Toolbox Standing Balance Test stances were used in this

proof-of-concept study, given the availability of detailed methods and reference data available [31,38,39] for comparison between devices.

Methods

Subjects

A total of 42 individuals participated in this study (Table 1). Procedures were approved by the Asentral, Inc. Institutional Review Board (Newburyport, MA, USA) and the US Army Human Research Protection Office. Subjects were recruited from the public. An informed consent form describing the nature of the study, as well as the exclusion criteria, was completed by all participants.

Participants were required to be between the ages of 18 and 39 years, weigh no more than 250 pounds, and possess normal hearing and normal or corrected-to-normal vision. Each of these criteria was confirmed by participant self-report. Participants were excluded if they reported any preexisting condition that may alter their ability to balance normally. A set of specific conditions that could affect balance were described for participants. Specific conditions listed for participants included multiple sclerosis, Parkinson's disease, Huntington's disease, other movement disorders, stroke, cervical spine or physical mobility issues, more than 1 fall in the past 6 months not as a result of an accident, current pregnancy, dizziness or vertigo, any lower extremity injury that required medical attention in the last 3 months, and any surgeries within the last year. All participants attested they were not taking any medication to lower blood pressure or to control a heart problem. All participants also attested they were not under instruction by a supervising physician to avoid full/unrestricted physical activity. Individuals were also screened based on self-report for history of a diagnosed seizure disorder (or any seizures within the last 3 years), as well as extreme sensory sensitivity. All participants also attested to having no diagnosed macular degeneration, glaucoma, or cataracts, or any chronic or acute conditions resulting in pain, including diabetes or a history of joint replacement.

Experimental Setup

Before administering the BAM protocol, subjects were outfitted with a gait belt. An Android smartphone (Samsung Galaxy S5, Samsung Galaxy S6, or LG Electronics/Google Nexus 5) was attached to the gait belt using a protective case with clip. The smartphone was attached upright, with the screen facing away from the subject. The subject was also given a pair of Glass Explorer Edition (henceforth, Glass) by the facilitator to wear. Subjects who normally wore glasses were given the option to wear Glass over or without their regular glasses. Subjects were asked to read a sentence on the display screen to confirm the screen was adjusted properly. A test exercise was administered on Glass to ensure subjects could (1) operate Glass by tapping on the side and (2) could hear a tone played from Glass.

The BAM protocol was administered as previously described [39]. The BAM protocol includes 6 standing conditions: (1) solid surface, feet together, eyes open, and (2) eyes closed; (3) foam surface (Airex Balance Pad, Specialty Foams, Switzerland), feet together, eyes open, and (4) eyes closed; (5) solid surface, tandem standing, eyes open, and (6) eyes closed (Figure 1). During each stance, all subjects were asked to stand quietly for 60 seconds and to look (in eyes-open conditions) at a symbol placed centrally at eye level 1 m from the subject. Subjects were instructed by the facilitator regarding stance following the instructions adapted from the NIH Toolbox Standing Balance Test [31,40]. Stance was also described on the smartglasses display screen. Subjects initiated each set of data collection by tapping the side of the smartglasses. A timer was displayed on Glass showing time remaining and a tone was played at the end of each timed stance. All subjects completed 2 attempts of all stances on the same day.

A trained study facilitator was present during the study and was ready to prevent the participant from falling. The study facilitator observed participants for failure to hold the demonstrated pose. Failures were recorded if (1) participant's arms came off his/her chest, (2) participant's knees bent, (3) participant's feet moved out of original position (move or swivel out or are lifted), (4) participant bent forward at the waist (more than 45°), (5) participant opened his/her eyes during an eye-closed pose, or (6) participant says something like "I cannot do that" or "I do not feel safe trying that."

Data Acquisition

An Android app was developed to synchronize recording of device IMU data between smartglasses and the waist-mounted smartphone (Figure 2). The app was installed on both Glass and the Android smartphones before testing. The app allowed Glass to pair with a smartphone via Bluetooth. Messages sent via Bluetooth from Google Glass to the smartphone were used to initiate a timer on Google Glass and begin storing IMU values (sampled at 50 Hz).

Table 1. Subject demographics (N=42).

Demographics	Mean (SD, range) or n (%)
Age in years, mean (SD, range)	23.8 (5.2, 18-37)
Gender, female, n (%)	16 (38)
Height in inches, mean (SD, range)	68 (3, 62-76)
Weight in pounds, mean (SD, range)	152 (32, 110-241)

Figure 1. Balance accelerometer measure (BAM) protocol conditions. (1, 2) Feet together on a firm surface used for conditions 1 (eyes open) and 2 (eyes closed). (3, 4) Feet together on a foam surface used for conditions 3 (eyes open) and 4 (eyes closed). (5, 6) Feet in tandem stance on a firm surface used for conditions 5 (eyes open) and 6 (eyes closed).

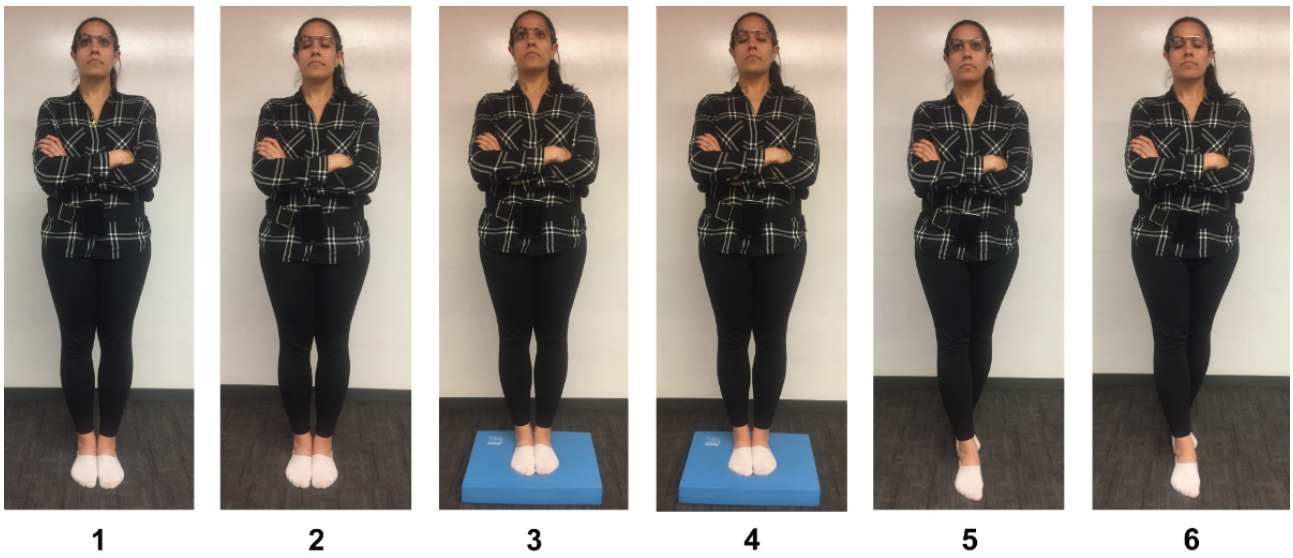
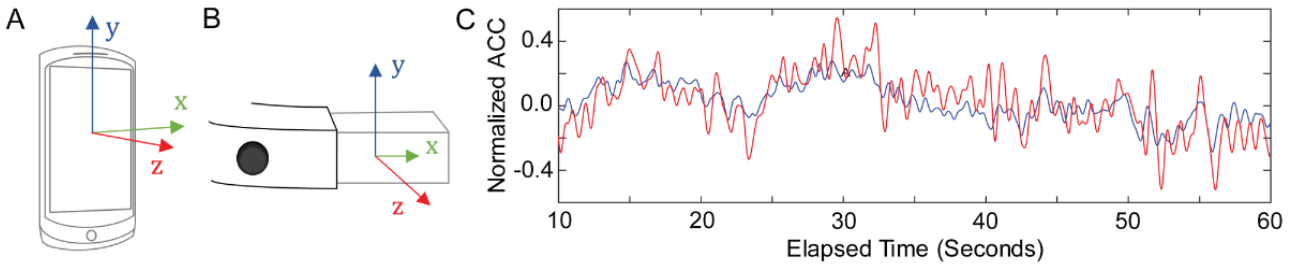


Figure 2. Accelerometer data (ACC) collection with smartphone and smartglasses. (A) Axes of accelerometer on Android smartphone compared with (B) Glass. (C) Example comparison of low-pass filtered ACC (z-axis) collected during a trial of condition 6 in Glass (red) compared with waist-mounted smartphone (blue).



Upon completion of each stance, time-stamped IMU data were saved on the device as a comma-delimited plain text file. When running on Google Glass, the app provides instructions on stance, a timer, and a tone that plays at the end of each stance session. At the end of each study session, data were transferred from respective devices to secure cloud storage.

Data Analysis

Data files were imported into MATLAB 2016b (MathWorks, Native, MA, USA) for analysis, which included use of the Signal Processing Toolbox, the Statistics and Machine Learning Toolbox, and custom scripts. The first 10 seconds of data were discarded to ensure stability of measures (50 seconds of data total). Accelerometer data (ACC) from each trial was low-pass filtered using a fourth-order, Butterworth filter with a cutoff frequency of 1.25 Hz [39]. The normalized path length (NPL; mG/sec; higher values indicate more sway) of the anterior-posterior (AP) postural sway data was calculated as previously described [39]. NPL was also calculated from the combined ACC magnitude.

Trials recorded as failures by the study facilitator were excluded from further analysis. Smartglasses-based measurements of NPL along different axes were compared with smartphone measurements using Spearman rank correlation coefficient [41].

For comparison of differences between stances, the nonparametric Kruskal-Wallis test was used to compare mean ranks [42,43]. Normality of measurements within stance conditions was evaluated by the Anderson-Darling test [44]. Significant differences between correlation coefficients were determined by treating them as Pearson coefficients and using the standard Fisher z-transformation to compare using a standard normal procedure [45]. Test-retest reliability of NPL measurements was estimated for each condition between the 2 sessions by calculating the 2-way random, single-measure intraclass correlation coefficient, $ICC_{2,1}$, and corresponding 95% CI [46,47]. NPL was standardized as previously described [39], and the composite score was calculated by adding together the standardized values across all 6 conditions.

Results

Measurement of Anterior-Posterior Sway with Smartglasses Correlates With Measurement at Waist

All 42 subjects successfully passed both trials on conditions 1 through 3, similar to previous reports [39]. Both trials of the eyes closed/foam surface condition (condition 4) were passed successfully by 37 subjects (88%). One subject failed a trial of the eyes-open/tandem stance condition (condition 5). Thirty

subjects (71%) successfully passed 2 trials of the eyes closed/tandem stance condition (condition 6). All observed failures were recorded as feet moving out of the original position and/or arms coming off the chest. Overall, 2 trials on all 6 conditions were successfully passed by 28 subjects (66%).

NPL AP sway measured from the head was strongly correlated (Spearman rank correlation coefficient=.85) with NPL AP sway measured from the waist (Figure 3). Mean NPL AP sway measured from the waist was in good agreement with previously reported values [39], although we observed a higher mean for condition 6. The mean (SD) composite score was 21.4 (18.0), which was in good agreement with the previously reported value of 19.6 (15.3) for healthy subjects.

Although NPL measured from the head was generally larger than NPL measured from the waist in each trial, mean NPL AP sway measured from the head in each condition was observed to follow a similar trend as the means measured from the waist. Significant differences (Kruskal-Wallis, $P<.001$) were found between each set of eyes-open and eyes-closed conditions as well as between standing on feet together/firm surface compared with foam surface or tandem stance.

Correlation Between Measurements Was Significantly Stronger When Calculating Normalized Path Length From All Three Axes

Measuring sway along the ACC's AP axis was previously shown to be sufficient to differentiate healthy subjects from subjects with vestibular disorders [39]. However, the additional ACC acquired from commercial off-the-shelf (COTS) smart devices may further enhance measurement accuracy, particularly along the mediolateral x-axis. Indeed, the NPL calculated using all 3 axes (total NPL) was found to have a significantly stronger correlation (Spearman rank correlation coefficient=.90, $P<.001$) between head- and waist-based measurements (Figure 4). Mean total NPL measured in each condition followed similar trends as using AP NPL only for both waist- and head-based measurements.

Test-Retest Reliability of Measures Were Comparable Between Head and Waist

Previously, the test-retest reliability of NPL AP measured from the waist was found to be generally good (ICC≥0.74) across all conditions, except for condition 6 [39]. Here, same-day test-retest reliability of AP NPL measured from the head with smartglasses was found to be very good (Figure 5), with an $ICC_{2,1}$ (95% CI) of 0.85 (0.81-0.88). This was comparable to our estimation of the test-retest reliability of waist-based AP NPL, which was 0.84 (0.80-0.88), agreeing with previously reported values.

Using total NPL, we found a slight improvement in test-retest reliability in both head- and waist-based measurements. $ICC_{2,1}$ (95% CI) was found to be 0.87 (0.83-0.90) in the case of head-based measurement, as opposed to 0.90 (0.88-0.92) in the case of waist-based measurement.

Figure 3. Anterior-posterior (AP) sway measured with smartglasses is highly correlated with waist-based accelerometry. (A) AP sway measured from the head was strongly correlated with AP sway measured from the waist (pooled data from all conditions with 95% prediction bands). Geometric mean and 95% CI for waist-based (B) and head-based (C) measurements of AP normalized path length (NPL) by condition.

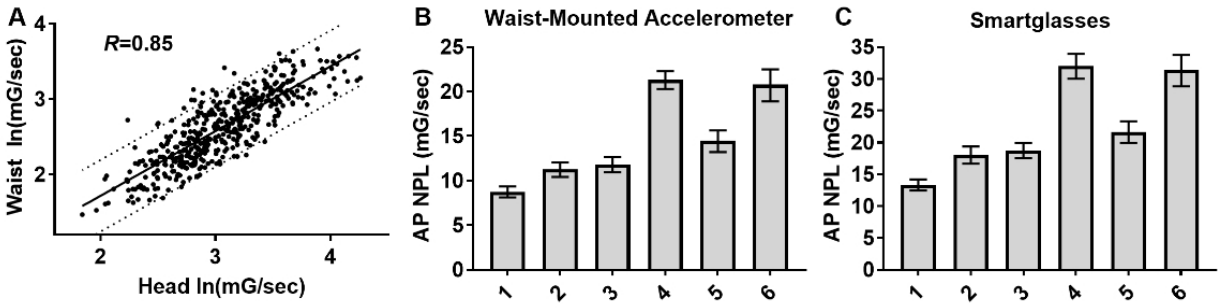


Figure 4. Total sway magnitude measured with smartglasses is highly correlated with waist-based accelerometry. (A) Total sway measured from head was more strongly correlated with sway measured from the waist. Geometric mean and 95% CI for waist-based (B) and head-based (C) measurement of total normalized path length (NPL) by condition.

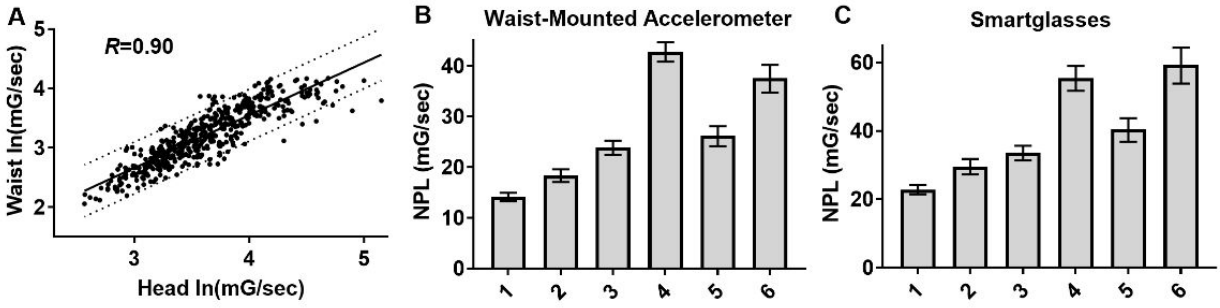
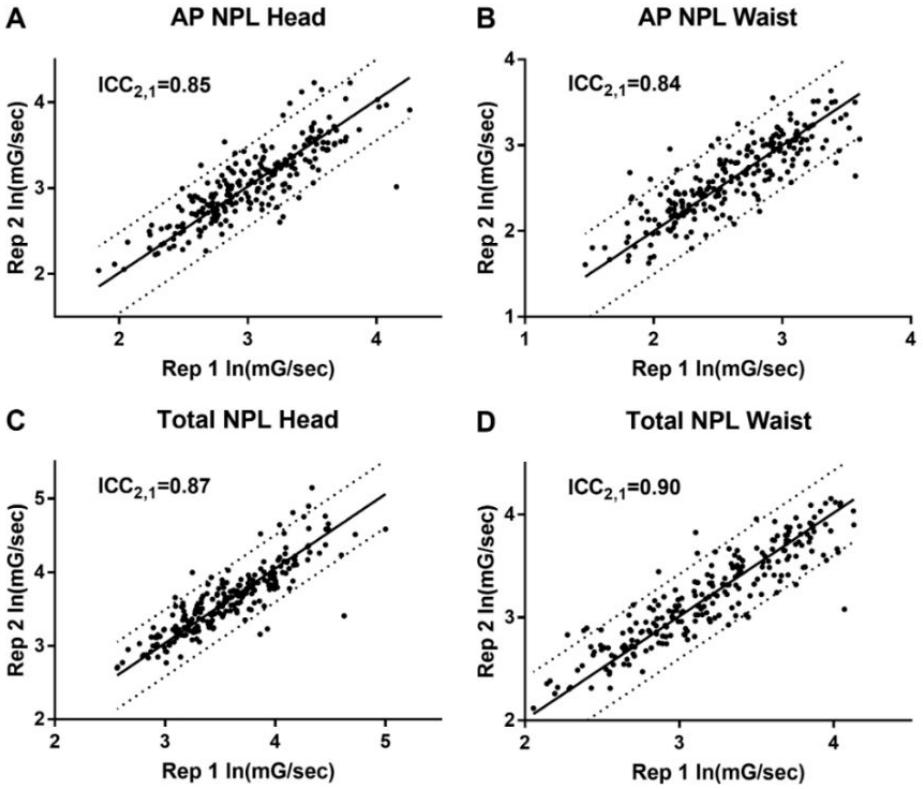


Figure 5. Head-mounted smartglasses have comparable test-retest reliability to waist-mounted balance accelerometer measurement. Test-retest intraclass correlation coefficient, ICC2,1, for accelerometer measures of postural sway along anterior-posterior (AP) (A=head, B=waist) and all axes (C=head, D=waist). NPL: normalized path length.



Discussion

Principal Findings

To the best of our knowledge, this is the first study assessing the feasibility of using COTS smartglasses to perform quantitative standing balance measurement. This study indicates that head-based measurement of AP or total sway with smartglasses following the BAM protocol produces similar results to waist-based measurement. This included similar relative differences between test conditions as well as similar test-retest reliability. Condition effects observed with this protocol previously supported the validity of waist AP NPL as a measure of balance. Similar to these previous findings, sway measured with smartglasses was larger with eyes closed than with eyes open for all stance conditions. Sway was also larger in tandem stance and on foam surface conditions compared with corresponding conditions with feet together on a firm surface. These condition effects previously indicated that ACC measured from the waist was sensitive to changes in the sensory modalities available for balance, including vision and somatosensation [48]. In this study, we demonstrate that head-based ACC measurement using COTS smartglasses has a comparable sensitivity for measuring these differences.

Comparison With Prior Work and Future Implications

The current wide availability of smartphones with IMU technology has made them an attractive platform to develop physical health assessments on. Along with standing balance, smartphone-based measurements are also being developed to objectively quantify a range of related functional mobility assessments [49-53]. Similarly, there are a growing number of dedicated wearables that have been developed to provide research- and clinical-grade balance, gait, full-body kinematics, and other functional mobility assessments. An important distinction between using the sensors in a smartphone compared with wearable hardware built for a specific function is that smartphones are already widely used by the population. Thus, smartphone-based assessments can be immediately accessible on this multifunctional platform. The multifunctional versatility of such a device could be particularly transformative in the battlefield, where there are practical limitations to amount of equipment that can be transported in various circumstances [28]. When considering the broader goal of having an objective balance assessment as only one component of a multifactorial concussion battery, neuroimaging and biomolecular assays could provide more definitive results and aid in differential diagnosis. However, the equipment needed to provide this level of certainty would be less practical for point-of-injury assessment and triage when a software-based assessment on a multipurpose device could sufficiently determine the need for additional care.

Leveraging the wide availability of COTS hardware to develop objective clinical assessments and rehabilitative strategies has motivated research into not just smartphones. The Nintendo Wii Balance Board [54-57] and Microsoft Kinect [58-60] are also being used in physical medicine. Indeed, there are now several FDA-cleared Kinect applications suitable for use in the clinic or at home that provide exercise guidance and remotely

accessible patient performance metrics [61,62]. The sophistication of sensor-rich COTS hardware enables health care apps to be developed without the costs typically associated with dedicated health care equipment design, manufacturing, quality control, storage, distribution, etc. Admittedly, smartglasses are far from reaching the ubiquity of these devices. However, the market for smartglasses is projected to reach 3.4 million units by 2020, with health care being a major driver of smartglasses' growth [63]. Smartglasses have been shown to be well tolerated in children and adults with autism spectrum disorder, providing evidence to support their use as an assistive device [64-66]. In the longer term, decline in costs, the solidification of applications and model features, and technology saturation of smartphone and tablet markets could push smartglasses to become a dominant consumer computing device [63]. It is this context, considering the potential future widespread availability of lightweight and portable head-mounted wearables, which motivates the research study described here.

Critically, it was not the objective of this study to determine whether smartglasses would provide a more sensitive measure than smartphones or act as a replacement for gold standard methods of clinical assessment. Rather, with this feasibility demonstrated, it can be discussed how smartglasses could have specific advantages over other COTS devices for assessment and rehabilitation of balance dysfunction related to brain injury. Recently, it was reported that a fully immersive head-mounted virtual reality (VR) system was successfully used to obtain repeatable balance assessment measurements in an elderly population [67]. Higher fall risk participants were found to change their tilt in the AP direction at a significantly higher rate. Although minimal simulator sickness was generally reported in this study, at least one participant dropped out of the study because of this issue. There were also significant differences in nausea pre- and postmeasurement. In terms of head-mounted wearables, smartglasses may be preferable to fully immersive VR headsets as they do not completely obscure external vision, which suggests they could be a safer alternative (Figure 6). Smartglasses could provide real-time feedback to correct balance instability during movement in an actual environment, such as through audio [68-70], and Glass has been shown to be feasible for external rhythmic cueing to improve gait in Parkinson's patients [71]. Furthermore, fully immersive VR headsets often include foam that is pressed against the user's face that can quickly become unsanitary, leading to hygiene concerns and the potential for disease transmission when used in a clinical setting [72]. Thus, smartglasses may be preferable in clinical use for not just balance assessment/feedback but also VR-based vestibular-ocular motor and cognitive assessments.

Limitations

This study is primarily a proof-of-concept demonstrating that measurements obtained from the IMU of a specific COTS head-mounted wearable (Glass) can provide quantitative balance measurements. These head-based measurements are comparable with waist-based measurements when following the NIH Standing Balance Toolbox protocol. This report only describes one quantitative balance measurement derived from accelerometry, NPL, although there are a variety of methods to preprocess these types of data [73]. A variety of subjective and objective assessments exist to both identify and characterize balance deficits [74]. A comprehensive characterization of measures obtainable from smartglasses against a clinical force plate system would provide a more thorough assessment of the concurrent validity of head-based measurement. Recently published pilot results from an elderly population using a force plate system support the potential for head-based measurement using COTS hardware in clinical assessment of balance [67].

Figure 6. Fully immersive head-mounted device compared with partially immersive smartglasses. Fully immersive virtual reality headsets, such as Oculus Rift (left), completely block external stimuli, limiting their use in concussed populations where dizziness, nausea, and sensitivity to light are common persistent symptoms. Smartglasses, such as Glass (right), could provide a safer, more portable, and lighter-weight alternative. Of note, Glass weighs only 1.3 ounces—more than 10 times less than Oculus Rift.

Stances in the BAM protocol and NIH Standing Balance Toolbox were used in this study as a standard for preliminary comparison, given the availability of detailed methods and baseline data [31,38,39]. However, it has been previously suggested that as this protocol was not designed with the goal of concussion assessment, it may have limited use in this domain in comparison with other protocols. In one report, BAM was found not to effectively discriminate between healthy and concussed adolescents. Rather, expert scoring of the Balance Error Scoring System (BESS) protocol was able to identify patients from healthy participants with 60% sensitivity and 82% specificity [48]. The BESS protocol is similar to the BAM protocol with 6 conditions in total. However, in BESS, all conditions are performed with eyes closed and hands positioned on the hip, with 3 stances (feet together, single leg, and tandem stance) performed on both firm and foam surfaces. A modified BESS protocol, which eliminates the foam surface conditions, has been included as part of the SCAT sideline concussion evaluation since the second edition [75]. Although the modified BESS protocol may lack sensitivity, instrumenting the modified BESS with a waist-based inertial sensor led to superior diagnostic classification of recently concussed individuals compared with BESS alone, albeit in a relatively small sample size of 13 recently concussed individuals and 13 demographically matched controls [76]. In summary, future evaluations should consider whether other procedures are necessary, depending on study goals.

The study presented here is limited by its exclusive use of healthy subjects. Further research is necessary to determine whether measurements using a head-mounted device can detect deficits in postural sway related to specific medical conditions. Although the BAM protocol was unable to sufficiently discriminate concussed adolescents, postural sway as measured by waist-based BAM using this protocol was able to discriminate between persons with peripheral vestibular impairments and those without balance impairment [39]. In general, postural sway measurement alone currently lacks the sensitivity and specificity needed to confidently diagnose concussions. It is important to reiterate that our study goal was to demonstrate the feasibility of obtaining objective and quantitative measurements of postural sway with a head-mounted wearable. We hypothesize this would serve as only one component of a concussion assessment battery that could be automatically administered using COTS smartglasses as a platform.

Finally, although we determined that head-based measurement was generally consistent with waist-based measurement, head-based measurement might present additional challenges when administered outside carefully monitored conditions. For example, head-based measurement may be more sensitive to behavioral artifacts such as speech and shifting attention. It is important to mitigate these challenges by detecting and removing these artifacts to improve the internal validity of the assessment when used independently for clinical decision-making.

Conclusions

The accelerometer built into Glass is sufficient to provide standing balance measurements comparable with commercial smartphones. Accelerometry measurements obtained from the head, including the NPL of AP sway as well as the total NPL magnitude, resulted in similar condition effects as those obtained from the waist in a healthy adult population. Head-based measurement of balance using smartglasses could serve as one component of a wearable, multifactorial concussion assessment that has integrated instruction and feedback. This approach could improve the objective assessment of concussion symptoms in high-risk activities, including contact sports and warfare, where current standards often rely on subjective methods, including symptom self-report by the injured. Smartglasses may provide a safer, lighter-weight, more portable, and more hygienic alternative to fully occlusive head-mounted wearables, while providing a similar range of assessments for concussion detection, including cognitive and vestibular-ocular motor screens. Further research is necessary to demonstrate the ability of smartglasses to detect balance dysfunction stemming from concussion.

Acknowledgments

This study is based upon work supported by the United States Army Medical Research and Materiel Command under Contract No. W81XWH-14-C-0007 (awarded to TIAX, LLC). Additional work was supported by the Office of the Assistant Secretary of Defense for Health Affairs, through the Autism Research Program under Award No. W81XWH-17-1-0449. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense. The authors thank Alyk Moomaw for her help in preparing the manuscript. The authors thank Google, Inc for generous support and the Glass team at X (formerly Google X) for technical guidance on Glass development.

Conflicts of Interest

NTS is the owner of Brain Power, LLC, and has patent rights associated with some subject matters discussed.

References

1. Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. J Head Trauma Rehabil 2006;21(5):375-378. [Medline: 16983222]
2. Defense and Veterans Brain Injury Center. 2016. DoD Worldwide Numbers for TBI URL: http://dvbic.dcoe.mil/dod-worldwide-numbers-tbi [accessed 2017-12-28] [WebCite Cache ID 6w3J66y9M]

3. Johnson MA, Hawley JS, Theeler BJ. Management of acute concussion in a deployed military setting. Curr Treat Options Neurol 2014 Sep;16(9):311. [doi: 10.1007/s11940-014-0311-8] [Medline: 25062543]
4. Wang Y, Chan RC, Deng Y. Examination of postconcussion-like symptoms in healthy university students: relationships to subjective and objective neuropsychological function performance. Arch Clin Neuropsychol 2006 May;21(4):339-347 [FREE Full text] [doi: 10.1016/j.acn.2006.03.006] [Medline: 16765018]
5. Landre N, Poppe CJ, Davis N, Schmaus B, Hobbs SE. Cognitive functioning and postconcussive symptoms in trauma patients with and without mild TBI. Arch Clin Neuropsychol 2006 May;21(4):255-273 [FREE Full text] [doi: 10.1016/j.acn.2005.12.007] [Medline: 16716563]
6. McCrea M, Iverson GL, McAllister TW, Hammeke TA, Powell MR, Barr WB, et al. An integrated review of recovery after mild traumatic brain injury (MTBI): implications for clinical management. Clin Neuropsychol 2009 Nov;23(8):1368-1390. [doi: 10.1080/13854040903074652] [Medline: 19882476]
7. Gaylord KM, Cooper DB, Mercado JM, Kennedy JE, Yoder LH, Holcomb JB. Incidence of posttraumatic stress disorder and mild traumatic brain injury in burned service members: preliminary report. J Trauma 2008 Feb;64(2 Suppl):S200-5; discussion S205-6. [doi: 10.1097/TA.0b013e318160ba42] [Medline: 18376167]
8. Kraus JF, McArthur DL. Incidence and prevalence of, and costs associated with, traumatic brain injury. In: Rosenthal M, editor. Rehabilitation of the Adult and Child With Traumatic Brain Injury. Philadelphia, PA: F A Davis Co; 1999:3-18.
9. Hoge CW, Toboni HE, Messer SC, Bell N, Amoroso P, Orman DT. The occupational burden of mental disorders in the U.S. military: psychiatric hospitalizations, involuntary separations, and disability. Am J Psychiatry 2005 Mar;162(3):585-591. [doi: 10.1176/appi.ajp.162.3.585] [Medline: 15741477]
10. Vasterling JJ, Proctor SP, Amoroso P, Kane R, Heeren T, White RF. Neuropsychological outcomes of army personnel following deployment to the Iraq war. J Am Med Assoc 2006 Aug 2;296(5):519-529. [doi: 10.1001/jama.296.5.519] [Medline: 16882958]
11. McCrea M, Hammeke T, Olsen G, Leo P, Guskiewicz K. Unreported concussion in high school football players: implications for prevention. Clin J Sport Med 2004 Jan;14(1):13-17. [Medline: 14712161]
12. Williamson IJ, Goodman D. Converging evidence for the under-reporting of concussions in youth ice hockey. Br J Sports Med 2006 Feb;40(2):128-32; discussion 128-32 [FREE Full text] [doi: 10.1136/bjsm.2005.021832] [Medline: 16431999]
13. Chapman JC, Diaz-Arrastia R. Military traumatic brain injury: a review. Alzheimers Dement 2014 Jun;10(3 Suppl):S97-104 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.012] [Medline: 24924680]
14. McCrory P, Meeuwisse W, Aubry M, Cantu B, Dvorak J, Echemendia R, et al. Consensus statement on Concussion in Sport--the 4th International Conference on Concussion in Sport held in Zurich, November 2012. J Sci Med Sport 2013 May;16(3):178-189 [FREE Full text] [doi: 10.1016/j.jsams.2013.02.009] [Medline: 23541595]
15. Coughlin SS. Post-traumatic stress disorder and traumatic brain injury. In: Post-Traumatic Stress Disorder and Chronic Health Conditions. Washington, DC: American Public Health Association; 2012.
16. Carroll L, Rosner D. The Concussion Crisis: Anatomy of a Silent Epidemic. New York, NY: Simon and Schuster; 2012.
17. Smith AM, Stuart MJ, Roberts WO, Dodick DW, Finnoff JT, Jorgensen JK, et al. Concussion in ice hockey: current gaps and future directions in an objective diagnosis. Clin J Sport Med 2017 Sep;27(5):503-509. [doi: 10.1097/JSM.0000000000000412] [Medline: 28079758]
18. Coldren RL, Kelly MP, Parish RV, Dretsch M, Russell ML. Evaluation of the military acute concussion evaluation for use in combat operations more than 12 hours after injury. Mil Med 2010 Jul;175(7):477-481. [Medline: 20684450]
19. Phillips T. The impact of litigation, regulation, and legislation on sport concussion management. Sport J 2015:8.
20. Abdullah KG, Grady MS, Levine JM. Concussion and football: a review and editorial. Curr Neurol Neurosci Rep 2015 Apr;15(4):11. [doi: 10.1007/s11910-015-0535-3] [Medline: 25702047]
21. Guskiewicz K, Teel E, McCrea M. Concussion: key stakeholders and multidisciplinary participation in making sports safe. Neurosurgery 2014 Oct;75 Suppl 4:S113-S118. [doi: 10.1227/NEU.0000000000000494] [Medline: 25232875]
22. Connelly C, Martin K, Elterman J, Zonies D. Early traumatic brain injury screen in 6594 inpatient combat casualties. Injury 2017 Jan;48(1):64-69. [doi: 10.1016/j.injury.2016.08.025] [Medline: 27639602]
23. Carlson KF, Kehle SM, Meis LA, Greer N, Macdonald R, Rutks I, et al. Prevalence, assessment, and treatment of mild traumatic brain injury and posttraumatic stress disorder: a systematic review of the evidence. J Head Trauma Rehabil 2011;26(2):103-115. [doi: 10.1097/HTR.0b013e3181e50ef1] [Medline: 20631631]
24. Ling GS, Ecklund JM. Traumatic brain injury in modern war. Curr Opin Anaesthesiol 2011 Apr;24(2):124-130. [doi: 10.1097/ACO.0b013e32834458da] [Medline: 21301332]
25. McCrory P, Meeuwisse W, Dvorak J, Aubry M, Bailes J, Broglio S, et al. Consensus statement on concussion in sport-the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med 2017 Jun;51(11):838-847. [doi: 10.1136/bjsports-2017-097699] [Medline: 28446457]
26. Lovell MR, Iverson GL, Collins MW, Podell K, Johnston KM, Pardini D, et al. Measurement of symptoms following sports-related concussion: reliability and normative data for the post-concussion scale. Appl Neuropsychol 2006;13(3):166-174. [doi: 10.1207/s15324826an1303_4] [Medline: 17361669]
27. Thompson JW, Hagedorn D. Multimodal analysis: new approaches to the concussion conundrum. J Clin Sport Psychol 2012 Mar;6(1):22-46. [doi: 10.1123/jcsp.6.1.22]

28. Schmid KE, Tortella FC. The diagnosis of traumatic brain injury on the battlefield. Front Neurol 2012 Jun 12;3:90 [FREE Full text] [doi: 10.3389/fneur.2012.00090] [Medline: 22701447]
29. Ozdalga E, Ozdalga A, Ahuja N. The smartphone in medicine: a review of current and potential use among physicians and students. J Med Internet Res 2012 Sep 27;14(5):e128 [FREE Full text] [doi: 10.2196/jmir.1994] [Medline: 23017375]
30. Shah N, Aleong R, So I. Novel use of a smartphone to measure standing balance. JMIR Rehabil Assist Technol 2016 Mar 29;3(1):e4 [FREE Full text] [doi: 10.2196/rehab.4511] [Medline: 28582247]
31. Rine RM, Schubert MC, Whitney SL, Roberts D, Redfern MS, Musolino MC, et al. Vestibular function assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S25-S31 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872c6a] [Medline: 23479540]
32. Patterson JA, Amick RZ, Thummar T, Rogers ME. Validation of measures from the smartphone sway balance application: a pilot study. Int J Sports Phys Ther 2014 Apr;9(2):135-139 [FREE Full text] [Medline: 24790774]
33. Amick RZ, Chaparro A, Patterson JA, Jorgensen MJ. Test-retest reliability of the SWAY Balance Mobile Application. J Mob Technol Med 2015 Jul 16;4(2):40-47. [doi: 10.7309/jmtm.4.2.6]
34. Amick RZ, Patterson JA, Jorgensen MJ. Sensitivity of tri-axial accelerometers within mobile consumer electronic devices: a pilot study. Int J Appl Sci Technol 2013;3(2):97-100 [FREE Full text] [doi: 10.13140/2.1.1568.8324]
35. Burghart M, Craig J, Radel J, Huisinga J. Reliability and validity of a mobile device application for use in sports-related concussion balance assessment. Curr Res Concussion 2017;4(1):e1-e6. [doi: 10.1055/s-0036-1597914]
36. Vincenzo JL, Glenn JM, Gray SM, Gray M. Balance measured by the sway balance smart-device application does not discriminate between older persons with and without a fall history. Aging Clin Exp Res 2016 Aug;28(4):679-686. [doi: 10.1007/s40520-015-0471-8] [Medline: 26458942]
37. Mitrasinovic S, Camacho E, Trivedi N, Logan J, Campbell C, Zilinyi R, et al. Clinical and surgical applications of smart glasses. Technol Health Care 2015;23(4):381-401. [doi: 10.3233/THC-150910] [Medline: 26409906]
38. Slotkin J, Kallen M, Griffith J, Magasi S, Salsman H, Nowinski C, et al. NIH Toolbox Standing Balance Test. In: NIH Toolbox Technical Manual. Bethesda, MD: National Institutes of Health; Sep 27, 2012.
39. Marchetti GF, Bellanca J, Whitney SL, Lin JC, Musolino MC, Furman GR, et al. The development of an accelerometer-based measure of human upright static anterior- posterior postural sway under various sensory conditions: test-retest reliability, scoring and preliminary validity of the Balance Accelerometry Measure (BAM). J Vestib Res 2013;23(4-5):227-235. [doi: 10.3233/VES-130490] [Medline: 24284603]
40. Reuben DB, Magasi S, McCreath HE, Bohannon RW, Wang YC, Bubela DJ, et al. Motor assessment using the NIH Toolbox. Neurology 2013 Mar 12;80(11 Suppl 3):S65-S75 [FREE Full text] [doi: 10.1212/WNL.0b013e3182872e01] [Medline: 23479547]
41. Zar JH. Spearman rank correlation. In: Armitage P, Colton T, editors. Encyclopedia of Biostatistics. Chichester, England: John Wiley & Sons; 1998.
42. Vargha A, Delaney HD. The Kruskal-Wallis test and stochastic homogeneity. J Educ Behav Stat 1998 Jan 1;23(2):170-192. [doi: 10.3102/10769986023002170]
43. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
44. Anderson TW, Darling DA. A test of goodness of fit. J Am Stat Assoc 1954;49:765-769. [doi: 10.2307/2281537]
45. Myers L, Sirois MJ. Spearman correlation coefficients, differences between. In: Encyclopedia of Statistical Sciences. New York, NY: Wiley; 2006.
46. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
47. Zoeller T. Mathworks. Intraclass correlation coefficient with confidence intervals URL: https://www.mathworks.com/matlabcentral/fileexchange/26885-intraclass-correlation-coefficient-with-confidence-intervals?requestedDomain=www.mathworks.com [accessed 2017-12-28] [WebCite Cache ID 6w3QWDKKU]
48. Furman GR, Lin CC, Bellanca JL, Marchetti GF, Collins MW, Whitney SL. Comparison of the balance accelerometer measure and balance error scoring system in adolescent concussions in sports. Am J Sports Med 2013 Jun;41(6):1404-1410 [FREE Full text] [doi: 10.1177/0363546513484446] [Medline: 23585486]
49. Madhushri P, Dzhagaryan AA, Jovanov E, Milenkovic A. A Smartphone Application Suite for Assessing Mobility. 2016 Presented at: Conf Proc IEEE Eng Med Biol Soc; 18 October 2016; Orlando, FL, USA. [doi: 10.1109/EMBC.2016.7591389]
50. Mellone S, Tacconi C, Chiari L. Validity of a smartphone-based instrumented Timed Up and Go. Gait Posture 2012 May;36(1):163-165. [doi: 10.1016/j.gaitpost.2012.02.006] [Medline: 22421189]
51. Lupton SD, Gimpel S, Jerzy R, Brunton LL, Hjerrild KA, Cosman D, et al. Characterization of the human and murine IL-7 genes. J Immunol 1990 May 1;144(9):3592-3601. [Medline: 2329282]
52. Madhushri P, Dzhagaryan A, Jovanov E, Milenkovic A. An mHealth tool suite for mobility assessment. Information 2016;7(3):47. [doi: 10.3390/info7030047]
53. Rehan YA, Gumaa M. Validity and reliability of smartphone applications for clinical assessment of the neuromusculoskeletal system. Expert Rev Med Devices 2017 Jun;14(6):481-493. [doi: 10.1080/17434440.2017.1325319] [Medline: 28462674]

54. Gil-Gómez JA, Lloréns R, Alcañiz M, Colomer C. Effectiveness of a Wii balance board-based system (eBaViR) for balance rehabilitation: a pilot randomized clinical trial in patients with acquired brain injury. J Neuroeng Rehabil 2011 May 23;8:30 [FREE Full text] [doi: 10.1186/1743-0003-8-30] [Medline: 21600066]
55. Goble DJ, Cone BL, Fling BW. Using the Wii Fit as a tool for balance assessment and neurorehabilitation: the first half decade of "Wii-search". J Neuroeng Rehabil 2014 Feb 8;11:12 [FREE Full text] [doi: 10.1186/1743-0003-11-12] [Medline: 24507245]
56. Prosperini L, Fortuna D, Giannì C, Leonardi L, Marchetti MR, Pozzilli C. Home-based balance training using the Wii balance board: a randomized, crossover pilot study in multiple sclerosis. Neurorehabil Neural Repair 2013;27(6):516-525. [doi: 10.1177/1545968313478484] [Medline: 23478168]
57. Wright WG, McDevitt J, Tierney R, Haran FJ, Appiah-Kubi KO, Dumont A. Assessing subacute mild traumatic brain injury with a portable virtual reality balance device. Disabil Rehabil 2017 Jul;39(15):1564-1572. [doi: 10.1080/09638288.2016.1226432] [Medline: 27718642]
58. Lange B, Chang CY, Suma E, Newman B, Rizzo AS, Bolas M. Development and evaluation of low cost game-based balance rehabilitation tool using the Microsoft Kinect sensor. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1831-1834. [doi: 10.1109/IEMBS.2011.6090521] [Medline: 22254685]
59. Paavola JM, Oliver KE, Ustinova KI. Use of X-box Kinect gaming console for rehabilitation of an individual with traumatic brain injury: a case report. J Nov Physiother 2013;3:129. [doi: 10.4172/2165-7025.1000129]
60. Venugopalan J, Cheng C, Stokes TH, Wang MD. Kinect-based rehabilitation system for patients with traumatic brain injury. Conf Proc IEEE Eng Med Biol Soc 2013;2013:4625-4628 [FREE Full text] [doi: 10.1109/EMBC.2013.6610578] [Medline: 24110765]
61. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. Nat Biotechnol 2016 Mar;34(3):239-246. [doi: 10.1038/nbt.3495] [Medline: 26963544]
62. Baumel A, Birnbaum ML, Sucala M. A systematic review and taxonomy of published quality criteria related to the evaluation of user-facing eHealth programs. J Med Syst 2017 Aug;41(8):128. [doi: 10.1007/s10916-017-0776-6] [Medline: 28735372]
63. Global Industry Analysts, Inc. Strategyr. Smart Augmented Reality (AR) Glasses - A Global Strategic Business Report URL: http://www.strategyr.com/Smart_Augmented_Reality_AR_Glasses_Market_Report.asp [accessed 2018-01-02] [WebCite Cache ID 6wBcQ3paa]
64. Keshav NU, Salisbury JP, Vahabzadeh A, Sahin NT. Social communication coaching smartglasses: well tolerated in a diverse sample of children and adults with autism. JMIR Mhealth Uhealth 2017 Sep 21;5(9):e140 [FREE Full text] [doi: 10.2196/mhealth.8534] [Medline: 28935618]
65. Liu R, Salisbury JP, Vahabzadeh A, Sahin NT. Feasibility of an autism-focused augmented reality smartglasses system for social communication and behavioral coaching. Front Pediatr 2017 Jun 26;5:145 [FREE Full text] [doi: 10.3389/fped.2017.00145] [Medline: 28695116]
66. Kinsella BG, Chow S, Kushki A. Evaluating the usability of a wearable social skills training technology for children with autism spectrum disorder. Front Robot AI 2017 Jul 12;4:31. [doi: 10.3389/frobt.2017.00031]
67. Saldana SJ, Marsh AP, Rejeski WJ, Haberl JK, Wu P, Rosenthal S, et al. Assessing balance through the use of a low-cost head-mounted display in older adults: a pilot study. Clin Interv Aging 2017 Aug 26;12:1363-1370 [FREE Full text] [doi: 10.2147/CIA.S141251] [Medline: 28883717]
68. Fino PC, Peterka RJ, Hullar TE, Murchison C, Horak FB, Chesnutt JC, et al. Assessment and rehabilitation of central sensory impairments for balance in mTBI using auditory biofeedback: a randomized clinical trial. BMC Neurol 2017 Feb 23;17(1):41 [FREE Full text] [doi: 10.1186/s12883-017-0812-7] [Medline: 28231824]
69. Dozza M, Horak FB, Chiari L. Auditory biofeedback substitutes for loss of sensory information in maintaining stance. Exp Brain Res 2007 Mar;178(1):37-48. [doi: 10.1007/s00221-006-0709-y] [Medline: 17021893]
70. Dozza M, Chiari L, Horak FB. Audio-biofeedback improves balance in patients with bilateral vestibular loss. Arch Phys Med Rehabil 2005 Jul;86(7):1401-1403. [Medline: 16003671]
71. Zhao Y, Nonnekes J, Storcken EJ, Janssen S, van Wegen EE, Bloem BR, et al. Feasibility of external rhythmic cueing with the Google Glass for improving gait in people with Parkinson's disease. J Neurol 2016 Jun;263(6):1156-1165 [FREE Full text] [doi: 10.1007/s00415-016-8115-2] [Medline: 27113598]
72. Martin-Lutz OH, Burmeister C, dos Santos LF, Morkisch N, Dohle C, Krüger J. Application of head-mounted devices with eye-tracking in virtual reality therapy. CDBME 2017;3(1):53-56 [FREE Full text] [doi: 10.1515/cdbme-2017-0012]
73. Mancini M, Salarian A, Carlson-Kuhta P, Zampieri C, King L, Chiari L, et al. ISway: a sensitive, valid and reliable measure of postural control. J Neuroeng Rehabil 2012 Aug 22;9:59 [FREE Full text] [doi: 10.1186/1743-0003-9-59] [Medline: 22913719]
74. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med 2010 Jun;46(2):239-248 [FREE Full text] [Medline: 20485226]
75. McCrory P, Meeuwisse W, Johnston K, Dvorak J, Aubry M, Molloy M, et al. Consensus Statement on Concussion in Sport: the 3rd International Conference on Concussion in Sport held in Zurich, November 2008. Br J Sports Med 2009 May;43 Suppl 1:i76-i90. [doi: 10.1136/bjsm.2009.058248] [Medline: 19433429]

76. King LA, Horak FB, Mancini M, Pierce D, Priest KC, Chesnutt J, et al. Instrumenting the balance error scoring system for use with patients reporting persistent balance problems after mild traumatic brain injury. Arch Phys Med Rehabil 2014 Feb;95(2):353-359. [doi: 10.1016/j.apmr.2013.10.015] [Medline: 24200875]

Abbreviations

ACC: accelerometer data
AP: anterior-posterior
BAM: balance accelerometer measure
BESS: Balance Error Scoring System
COTS: commercial off-the-shelf
FDA: Food and Drug Administration
ICC: intraclass correlation coefficient
IMU: inertial measurement unit
MACE: Military Acute Concussion Evaluation
mTBI: mild traumatic brain injury
NIH: National Institutes of Health
NPL: normalized path length
SCAT: Standardized Concussion Assessment Tool
VR: virtual reality

Edited by C Dias; submitted 14.07.17; peer-reviewed by N Shah, I Cikajlo, J Rey-Martinez, E Stanmore; comments to author 24.08.17; revised version received 14.10.17; accepted 21.11.17; published 23.01.18.

Please cite as:
Salisbury JP, Keshav NU, Sossong AD, Sahin NT
Concussion Assessment With Smartglasses: Validation Study of Balance Measurement Toward a Lightweight, Multimodal, Field-Ready Platform
JMIR Mhealth Uhealth 2018;6(1):e15
URL: http://mhealth.jmir.org/2018/1/e15/
doi:10.2196/mhealth.8478
PMID:29362210

Original Paper

Physical Activity Assessment Using an Activity Tracker in Patients with Rheumatoid Arthritis and Axial Spondyloarthritis: Prospective Observational Study

Charlotte Jacquemin[1,2], MD; Hervé Servy[3], MSc; Anna Molto[4,5], MD; Jérémie Sellam[6,7], PhD; Violaine Foltz[1,2], MD; Frédérique Gandjbakhch[1,2], MD; Christophe Hudry[4], MD; Stéphane Mitrovic[1,2], MD; Bruno Fautrel[1,2], PhD; Laure Gossec[1,2], PhD

[1]Rheumatology Department, Pitié Salpêtrière Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

[2]GRC-UPMC 08 (EEMOIS), UPMC Univ Paris 06, Sorbonne Université, Paris, France

[3]Sanoïa, E-Health Services, Gardanne, France

[4]Rheumatology B Department, Cochin Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

[5]INSERM (U1153), Clinical Epidemiology and Biostatistics, Paris-Descartes University, Sorbonne Paris-Cité, Paris, France

[6]Rheumatology Department, St-Antoine Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

[7]DHU i2B, INSERM UMRS_938, UPMC Univ Paris 06, Sorbonne Université, Paris, France

Corresponding Author:
Charlotte Jacquemin, MD
Rheumatology Department
Pitié Salpêtrière Hospital
Assistance Publique-Hôpitaux de Paris
47-83 bd de l'hôpital
Paris, 75013
France
Phone: 33 688624123
Email: jacquemin.charlotte@gmail.com

Abstract

Background: Physical activity can be tracked using mobile devices and is recommended in rheumatoid arthritis (RA) and axial spondyloarthritis (axSpA) management. The World Health Organization (WHO) recommends at least 150 min per week of moderate to vigorous physical activity (MVPA).

Objective: The objectives of this study were to assess and compare physical activity and its patterns in patients with RA and axSpA using an activity tracker and to assess the feasibility of mobile devices in this population.

Methods: This multicentric prospective observational study (ActConnect) included patients who had definite RA or axSpA, and a smartphone. Physical activity was assessed over 3 months using a mobile activity tracker, recording the number of steps per minute. The number of patients reaching the WHO recommendations was calculated. RA and axSpA were compared, using linear mixed models, for number of steps, proportion of morning steps, duration of total activity, and MVPA. Physical activity trajectories were identified using the K-means method, and factors related to the low activity trajectory were explored by logistic regression. Acceptability was assessed by the mean number of days the tracker was worn over the 3 months (ie, adherence), the percentage of wearing time, and by an acceptability questionnaire.

Results: A total of 157 patients (83 RA and 74 axSpA) were analyzed; 36.3% (57/157) patients were males, and their mean age was 46 (standard deviation [SD] 12) years and mean disease duration was 11 (SD 9) years. RA and axSpA patients had similar physical activity levels of 16 (SD 11) and 15 (SD 12) min per day of MVPA (P=.80), respectively. Only 27.4% (43/157) patients reached the recommendations with a mean MVPA of 106 (SD 77) min per week. The following three trajectories were identified with constant activity: low (54.1% [85/157] of patients), moderate (42.7% [67/157] of patients), and high (3.2% [5/157] of patients) levels of MVPA. A higher body mass index was significantly related to less physical activity (odds ratio 1.12, 95% CI 1.11-1.14). The activity trackers were worn during a mean of 79 (SD 17) days over the 90 days follow-up. Overall, patients considered the use of the tracker very acceptable, with a mean score of 8 out 10.

Conclusions: Patients with RA and axSpA performed insufficient physical activity with similar levels in both groups, despite the differences between the 2 diseases. Activity trackers allow longitudinal assessment of physical activity in these patients. The good adherence to this study and the good acceptability of wearing activity trackers confirmed the feasibility of the use of a mobile activity tracker in patients with rheumatic diseases.

(JMIR Mhealth Uhealth 2018;6(1):e1) doi:10.2196/mhealth.7948

KEYWORDS

fitness tracker; exercise; rheumatoid arthritis; axial spondylarthritis

Introduction

Physical Activity Recommendations

Recommendations for management of inflammatory joint diseases, in particular rheumatoid arthritis (RA) and axial spondyloarthritis (axSpA), state that physical activity should be encouraged [1,2]. Physical activity is not only important for general health [3,4] but it also helps to reduce pain and improve quality of life in rheumatic diseases [5,6]. To meet the World Health Organization (WHO) guidelines, a minimum of 150 min per week of moderate activity or 75 min per week of vigorous activity is recommended [7]. These guidelines have been converted in terms of steps per day and correspond, on average, to 7000 to 11,000 steps per day [8]. The threshold of 10,000 steps per day is usually retained for healthy persons [8-10]. The intensity of physical activity may be estimated by the step cadence (number of steps per minute) as follows: 100 and 130 steps per min correspond approximately to moderate and vigorous physical activity, respectively, and a cadence below 20 steps per min is considered as rest [8,9].

Assessment of Physical Activity in RA and axSpA

From previous published studies, patients with RA and axSpA appear to have low physical activity levels compared with guidelines and to healthy controls [11-16]. For example, a recent study found a median of 3710 steps per day in patients with RA based on a pedometer assessment [15]. However, these studies were cross-sectional and did not use mobile devices. Mobile activity trackers, such as smartwatches or "connected bracelets" (Withings, FitBit, Jawbone, and MisFit), allow both an interactive feedback on physical activity and the visualization of the evolution of precised activity patterns over time [17]. There is a growing interest in their use and their place in the management of chronic conditions [18,19], including in the field of rheumatology for noninflammatory diseases [20]. In patients with inflammatory joint disease, little is known regarding patterns of physical activity. Mobile activity trackers may contribute to determine profiles of patients with inflammatory joint disease for different physical activity patterns [16], according to duration, intensity, and frequency of physical activity.

Objectives

The primary objectives of this study were to measure and compare physical activity using a mobile activity tracker over 3 months, between RA and axSpA according to different physical activity patterns, including the number of steps and the duration of moderate to vigorous activity; to assess the proportion of patients reaching the recommendations; to determine trajectories of patients with different evolutions of physical activity; and to explore factors associated with a low physical activity. A secondary objective was to assess the feasibility of such a study using a mobile device.

Methods

Study Design

The ActConnect study was a prospective, multicenter, pragmatic, longitudinal observational study. It took place in 6 participating centers (3 tertiary care hospitals and 3 private practice physicians' offices) in Paris, France. All patients received full information at inclusion and provided informed consent. Ethical approval was obtained from the institutional review board (CPP Ile de France VI) and the human research ethics committee (CCTIRS, number 16.057bis).

Patients and Controls

Patients were eligible if they were above 18 years of age; had definite RA or axSpA according to the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria [21] or to the Assessment of SpondyloArthritis international Society (ASAS) [22] classification criteria for RA and axSpA, respectively; had an Internet access; and if they owned a mobile phone or tablet, which was compatible with the mobile activity tracker. All consecutive patients who satisfied the inclusion criteria, seen as outpatients (either consultation or day care hospital) by one of the investigators in the participating centers, were asked to participate between January 4 and April 29, 2016.

For indicative purposes, 20 healthy controls, with no rheumatic disease and aiming for similar mean age and gender distribution as for patients, were recruited from a convenience sample.

Data Collection

Medical Data

Characteristics of patients with RA and axSpA were collected at baseline and included sex, age, type of arthritis, disease duration (from diagnosis), and ongoing pharmacological arthritis treatment (including biologics and stability of the treatment during the last 3 months).

Comorbidities were collected using the Functional Comorbidity Index, which ranges from 0 (0=no comorbidity; however, the minimal score was 1 in this study because of the rheumatic disease) to 18 [23]. Where available, in patients with RA, the status for rheumatoid factor (RF) and for anticyclic citrullinated peptide (anti-CCP) and the presence of radiographic erosions were recorded; in patients with axSpA, the Human Leukocyte

Antigen (HLA) B27 status, history of peripheral and of extra-articular symptoms, and the presence of sacroiliitis on x-ray or on magnetic resonance imaging (MRI) were recorded. All imaging data were collected as recorded in the medical files, based on local readings in the context of usual care. Disease activity was assessed by the last available disease activity score 28 (DAS28) in patients with RA and by the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) on a 0 to 10 scale in patients with axSpA [24-26]; all patients filled in a patient global assessment (PtGA) [27]. Disability was self-assessed at the end of the 3 months using the modified Health Assessment Questionnaire (mHAQ), which ranges between 0 and 3 [28], and using the Bath Ankylosing Spondylitis Functional Index (BASFI) in patients with axSpA.

General Data

Patients and controls self-reported weight, height, socioprofessional category, work status, and current use of an activity tracker.

Physical Activity Data Collection

Each participant received an activity tracker (Withings Activité Pop [29]) and was instructed to wear it every day for 3 months. The Withings tracker is a watch with analog time display, and it records the number of steps per minute. This device was selected for practical aspects (months of autonomy and waterproof) and precision (steps per minute). The watch needed to be connected by Bluetooth, at least every 2 days, to a mobile phone app, which automatically transferred the physical activity data to the database. Data for 90 consecutive days from the first Monday following activation of the device were collected. No intervention was specifically performed to increase physical activity, and no instruction about physical activity was given to the participants.

Definitions and Description of Physical Activity Patterns

Physical activity was assessed by the number of steps per day (from 00:00 h to 23:59 h). Different physical activity patterns were also considered: proportion of morning steps over a day (number of steps between 00:00 h and 12:00 h, over the total number of daily steps) to reflect morning stiffness and activity duration in moderate to vigorous intensity (sum over a day of minutes with at least 100 steps recorded). The total activity duration (sum of minutes with at least 20 steps per min) and the number of steps per day in moderate to vigorous intensity (at least 100 steps per min) were also assessed but were strongly correlated to number of steps per day and to activity duration in moderate to vigorous intensity, respectively. The proportion of patients reaching the WHO recommendations [7] in terms of duration of moderate to vigorous activity per week and according to the threshold of 10,000 steps per day was calculated.

Acceptability Questionnaire

At the end of the 3 months, all patients answered an acceptability questionnaire prepared for this study. Ten questions were selected based on main barriers that emerged after interviewing 5 patients with RA and 5 patients with axSpA. Questions included, among others, are as follows: difficulties or discomfort to handle the bracelet because of the arthritis, worries about security of the bracelet data, and perceived utility of the activity tracker in daily life (Multimedia Appendix 1).

Statistical Analysis

Sample Size

To detect a difference of 1500 steps per day between patients with RA and axSpA, with a standard deviation (SD) of 3300 [15], and considering a statistical power of 80% and a significance level of 5%, 76 participants had to be analyzable in each group. For indicative purposes, 20 healthy controls were also included as anchor comparators.

Study Population

Patients and controls were analyzed only if they wore the bracelet for at least 60 complete days (of the 90). Only “full” days were analyzed (ie, at least 8 h recorded by the tracker between the first and the last steps). For sensitivity analyses, each missing day was imputed for the different activity patterns by the mean of the same weekdays for which data were available.

Physical Activity Over 3 Months

Physical activity patterns and their distributions over time were described in patients and controls and were compared between RA and axSpA using linear mixed-effect models. The nlme package in R was used [30]. Models included random intercept and slopes for patients and fixed effect for type of arthritis. The timepoints were each day from day 1 to day 90. When observed, the heteroscedasticity was modeled using the power variance function. Covariates included the following baseline characteristics, which visually differed between groups and/or may impact physical activity: sex, age, body mass index (BMI), disease duration, biologics, employment status, and PtGA. The probability for RA and axSpA patients to reach the WHO recommendations according to their rheumatic diagnosis was assessed by logistic regression adjusted on the same covariates.

Trajectories of Physical Activity

We assumed that patients may have different evolution of their physical activity over 3 months; some patients may tend to increase or decrease their physical activity because of the motivational aspect of the activity tracker or because of loss of this motivation. For each physical activity pattern, patients’ trajectories were partitioned using the K-means method adapted to longitudinal data in the R package Kml [31,32]: three clusters of patients with homogeneous trajectories of physical activity were identified. The K-means method is an explanatory analysis and needs no assumption about trajectories before running the algorithm. Different quality criteria allow to select the best partition, that is, the best number of clusters, based on the highest between-cluster covariance (well-separated clusters) and on the lowest within-cluster covariance (compact clusters). To explore factors associated with the low physical activity cluster, a multivariable logistic regression was performed. Covariates included type of arthritis, sex, age, BMI, disease duration, biologics, employment status, and baseline PtGA.

Acceptability and Adherence to the Activity Tracker

Acceptability was assessed in all included patients by the mean number of days the tracker was worn over the 3 months (ie, adherence), the percentage of wearing time, and by an acceptability questionnaire.

All analyses were performed using R version 3.2.2 [33].

Results

Demographic Characteristics

Overall, 178 patients and 20 controls were included. Among them, 157 patients (83 patients with RA and 74 patients with axSpA), with a total of 13,179 days of recording, and 19 controls wore the bracelet for at least 60 complete days and were analyzed (Figure 1). Patients had a mean age of 45.8 (SD 12.5) years, a mean BMI of 25.1 (SD 4.5) kg/m^2, and a mean disease duration of 10.5 (SD 8.9) years (Table 1). The majority (76.4%, 120/157) of patients were working. Patients with RA were mostly females and older than patients with axSpA; 61% (51/83) had radiographic erosions, and 80% (63/79) had positive RF and/or anti-CCP. In patients with axSpA, 59% (41/70) had past or present peripheral symptoms, 41% (30/73) had experienced extraarticular symptoms, 74% (49/66) were HLA B27 positive, and 82% (55/67) had radiographic and/or MRI sacroiliitis. Disease activity was well controlled (mean PtGA: 3.3 [SD 2.4]; mean DAS28 in RA: 2.3 [SD 1.2]; mean BASDAI in axSpA: 3.2 [SD 2.1]), and many patients were treated with biologics. Patients were little disabled; in RA, mean mHAQ was 0.2 (SD 0.4), and in axSpA, mean mHAQ was 0.3 (SD 0.3) and mean BASFI was 1.7 (SD 1.8). Treatments were globally stable; 19.1% (30/157) patients reported a change over the 3 months in any treatment specific to their arthritis. Controls' demographic characteristics visually appeared similar to patients (Table 1). The 21 patients not included in the analysis had comparable characteristics to those included (data not shown).

Physical Activity Over 3 Months

Patients performed, on average, 7124 (SD 2316) steps per day, corresponding to 108 (SD 36) min per day of activity, of which 16 (SD 11) min per day (ie, 106 [SD 77] min per week) corresponded to moderate to vigorous intensity (Table 2; Figure 2; Figure 3). Controls tended to be more active than patients with, on average, 9153 (SD 4127) steps per day, corresponding to 132 (SD 60) min per day of activity of which 26 (SD 20) min per day (ie, 174 [SD138] min per week) corresponded to moderate to vigorous intensity (Table 2; Figure 2). Thus, patients had 20% less total activity duration and daily steps compared with controls. This difference increased to 40% for moderate to vigorous activity duration. Overall, 27.4% (43/157) patients reached the WHO recommendations (with a mean of 204 [SD 56] min per week). Fifteen out of 157 (9.6%) patients walked with a mean of more than 10,000 steps per day (Table 2). Eight out of 19 (42%) and 5 out of 19 (26%) controls reached the following recommandations, respectively: mean 150 min per week of moderate to vigorous activity and mean 10,000 steps per day.

In longitudinal analyses, there was no statistically significant difference between patients with RA and axSpA in physical activity levels or patterns (Table 2).

The different activity patterns varied across days with periodic weekly variations in patients (Figures 4 and 5) and in controls (data not shown). Sundays seemed to be related to a decrease in all activity patterns (note that days 0, 7, 14, etc represent Sundays in Figures 4 and 5). The mean level of physical activity appeared to remain constant over the 3 months in patients and controls.

Figure 1. Flowchart of 178 included RA and axSpA patients. RA: rheumatoid arthritis; axSpA: axial spondyloarthritis.

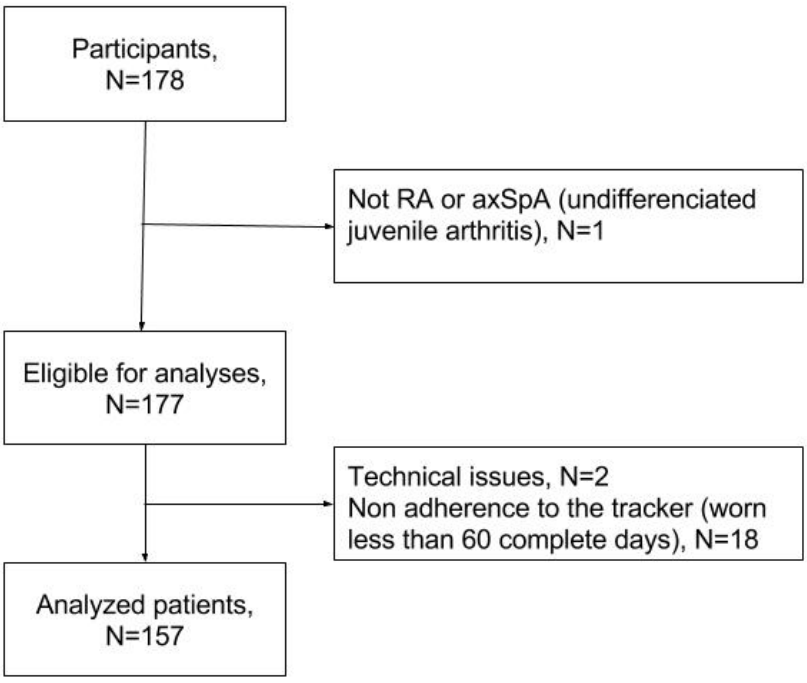


Table 1. Characteristics of 157 patients and 19 controls participating in an observational study of physical activity using an activity tracker.

Characteristics	RA[a] (N=83)	axSpA[b] (N=74)	Controls (N=19)
Sex, males, n (%)	14 (16.9)	43 (58.1)	8 (42.1)
Age in years, mean (SD[c])	49.9 (12.9)	41.3 (10.4)	45.1 (11.4)
BMI[d], kg/m^2, mean (SD)	24.9 (4.4)	25.3 (4.6)	24.2 (4.0)
Disease duration in years, mean (SD)	10.7 (8.8)	10.4 (9.1)	—
Work status, employed, n (%)	59 (71.1)	61 (82.4)	18 (94.7)
Manual work	3 (5.1)	3 (4.9)	0 (0.0)
Intellectual work	56 (94.9)	58 (95.1)	18 (100.0)
University studies, n (%)	68 (81.9)	62 (83.8)	17 (89)
Physical activity record history, n (%)	14 (16.9)	15 (20.3)	7 (36.8)
Functional comorbidity Index (ranging from 0 to 18), mean (SD)	1.6 (0.9)	1.3 (0.9)	
mHAQ[e], mean (SD)	0.2 (0.4)	0.3 (0.3)	
PtGA[f], mean (SD)	3.1 (2.3)	3.7 (2.5)	
Treatment at inclusion NSAIDs[g], n (%)	18 (21.7)	47 (63.5)	
Oral glucocorticoids, n (%)	21 (25.3)	1 (1.4)	
Synthetic DMARDs [h] , n (%)	78 (94.0)	17 (23.0)	
Methotrexate	68 (87.2)	13 (76.5)	
Biological therapy, n (%)	42 (50.6)	46 (62.2)	
antiTNF[i]	27 (64.3)	46 (100.0)	

[a]RA: rheumatoid arthritis.
[b]axSpA: axial spondyloarthritis.
[c]SD: standard deviation.
[d]BMI: body mass index.
[e]mHAQ: modified Health Assessment Questionnaire.
[f]PtGA: patient global assessment.
[g]NSAIDs: nonsteroidal antiinflammatory drugs.
[h]DMARDs: disease modifying antirheumatic drugs.
[i]antiTNF: antitumor necrosis factor.

Table 2. Physical activity patterns in RA and axSpA patients.

Physical activity patterns	RA[a] (N=83)	axSpA[b] (N=74)	*P* value[c]	Controls (N=19)
Total activity duration (min/day)[d], mean (SD[e])	108 (39)	108 (31)	.51[f]	132 (60)
Activity duration/day in moderate to vigorous intensity (min/day)[g], mean (SD)	16 (11)	15 (12)	.80[f]	26 (20)
Number of steps/day, mean (SD)	7118 (2411)	7130 (2221)	.50[f]	9153 (4127)
Percentage of steps before 12 AM, mean (SD)	32 (8)	32 (7)	.53[f]	36 (6)
Participants fulfilling the WHO[h] recommendation (mean physical activity≥150 min/week of moderate to vigorous activity), n (%)	25 (30)	18 (24)	.76[i]	8 (42)
Participants fulfilling the mean physical activity of ≥10,000 steps/day recommendations, n (%)	9 (11)	6 (8)	.71[i]	5 (26)

[a]RA: rheumatoid arthritis.

[b]axSpA: axial spondyloarthritis.

[c]*P* value of the comparison between RA and axSpA patients.

[d]At least 20 steps per min.

[e]SD: standard deviation.

[f]Comparison between RA and axSpA using linear mixed models adjusted on baseline characteristics.

[g]At least 100 steps per min.

[h]WHO: World Health Organization.

[i]Comparison between RA and axSpA using logistic regression adjusted on baseline characteristics.

Figure 2. Distribution of physical activity patterns in 83 rheumatoid arthritis and 74 axial spondyloarthritis patients and in 19 controls: (A) mean duration of moderate to vigorous activity (min per day) and (B) mean number of daily steps (steps per day). RA: rheumatoid arthritis; axSpA: axial spondyloarthritis.

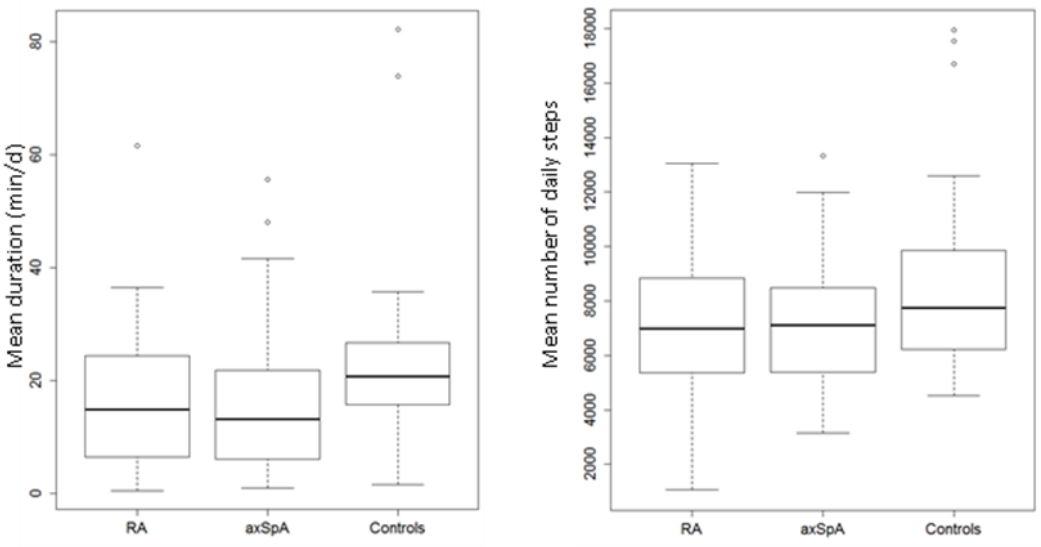


Figure 3. Distribution of mean physical activity over 3 months in 83 RA, 74 axSpA and 19 controls: (A) mean duration of moderate to vigorous activity (min/d) and (B) mean number of steps per day. RA: rheumatoid arthritis; axSpA: axial spondyloarthritis.

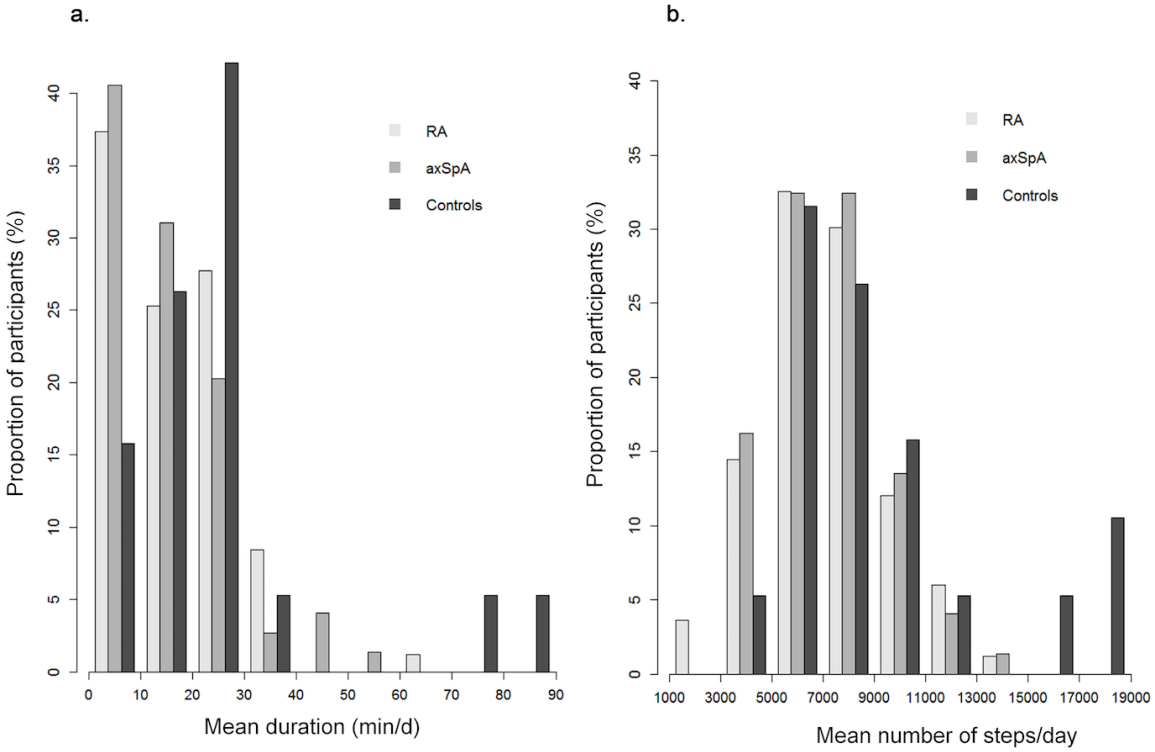


Figure 4. Weekly fluctuations of physical activity in 83 rheumatoid arthritis patients over 90 days, according to: (A) moderate to intense activity duration and (B) number of steps.

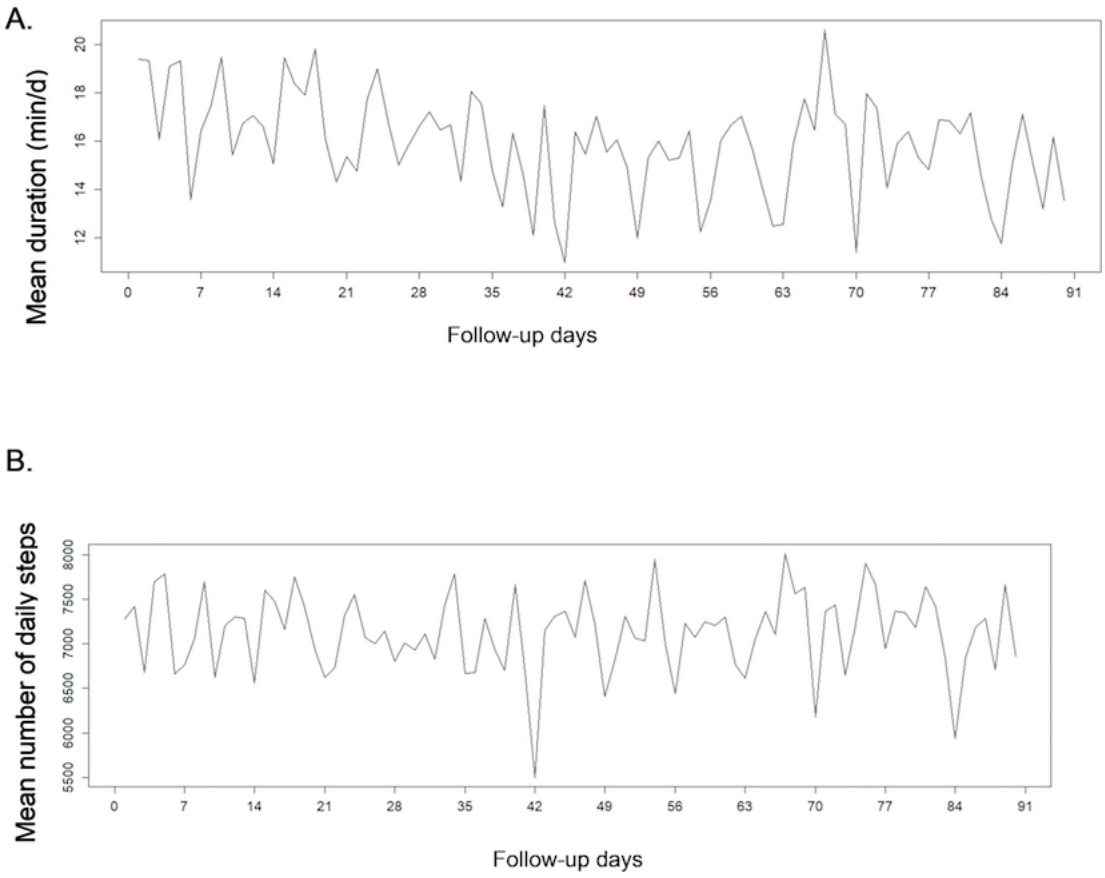
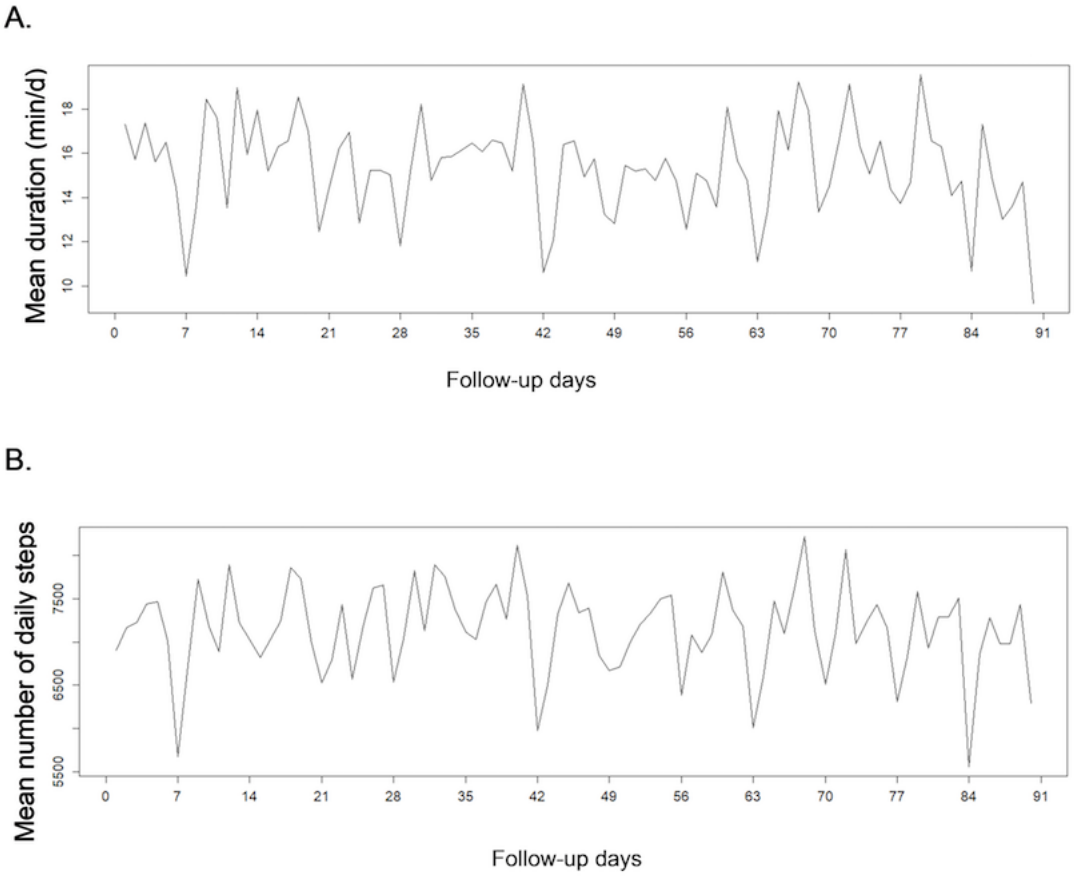


Figure 5. Weekly fluctuations of physical activity in 74 axialspondyloarthritis patients over 90 days, according to: (A) moderate to intense activity duration and (B) number of steps.



Comparisons of Clusters of Patients According to Their Trajectories

Patients were partitioned in 3 clusters with homogeneous trajectories according to following 3 levels of activity: low, moderate, and high (Figure 6). In all clusters, trajectories were overall constant over time.

Considering the duration of moderate to vigorous activity (Table 3), 54.1% (85/157) patients had a low activity level (mean of 7.2 [SD 4.2] min per day), 42.7% (67/157) had a moderate activity level (mean of 23.8 [SD 5.3] min per day), and 3.2% (5/157) had a high activity level (mean of 49.4 [SD 9.2] min per day). Patients with RA in the low activity cluster (n=44) had a mean DAS 28 of 2.1 (SD 0.9) versus 2.6 (SD 1.4) in the moderate to high cluster (n=39). Patients with axSpA had, respectively, a BASDAI of 2.7 (SD 2.1) and 3.6 (SD 2.0) in the low activity cluster (n=41) and the moderate to high activity cluster (N=33). The main factor associated with being in the low activity cluster was a higher BMI (odds ratio [OR] for a 5-point increase, OR 1.13, CI 1.11-1.15, *P*=.007).

Considering the number of steps per day (Table 3), 38.2% (60/157) patients were in the low activity cluster (mean of 4823 [SD 1020] steps per day), 46.5% (73/157) were in the moderate activity cluster (mean of 7789 [SD 877] steps per day), and 15.3% (24/157) were in the high activity cluster (mean of 10,852 [SD 1259] steps per day). BMI and biological therapy were associated with the low activity cluster.

In sensitivity analyses, all results were confirmed after imputation of missing days (data not shown).

Acceptability of the Activity Tracker

For the 177 eligible patients (Figure 1), the activity trackers were worn during a mean of 79 (SD 17) days, corresponding to a mean of 88% (SD 19) of days over the period; 70.6% (125/177) patients wore it for at least 80 out of 90 days, and 78.5% (139/177) patients still wore the device at the end of the 3-month period. Patients (N=171) considered the use of the watch very acceptable, with a mean score of 8 out 10, and self-reported barriers were rare (Multimedia Appendix 1). At the end of the study, 63.2% (108/171) patients were considering to keep wearing the tracker most of the time. Only 4.5% (8/177) patients (4 with RA and 4 with axSpA) reported to have problem with the watch because of their arthritis. When the adherent patients (ie, analyzed ones, N=153) were compared with the nonadherent patients (N=18) for acceptability, adherent patients tended to find the tracker slightly more acceptable; the mean acceptability scores were 8.6 out 10 (SD 2.4) and 7.8 out 10 (SD 2.8), respectively; and 65% versus 50% of adherent versus nonadherent patients considered to keep wearing the tracker regularly after the end of the study.

Figure 6. Mean trajectories over 3 months of all patients partitioned in 3 clusters according to: (A) moderate to vigorous activity duration and (B) number of steps per day.

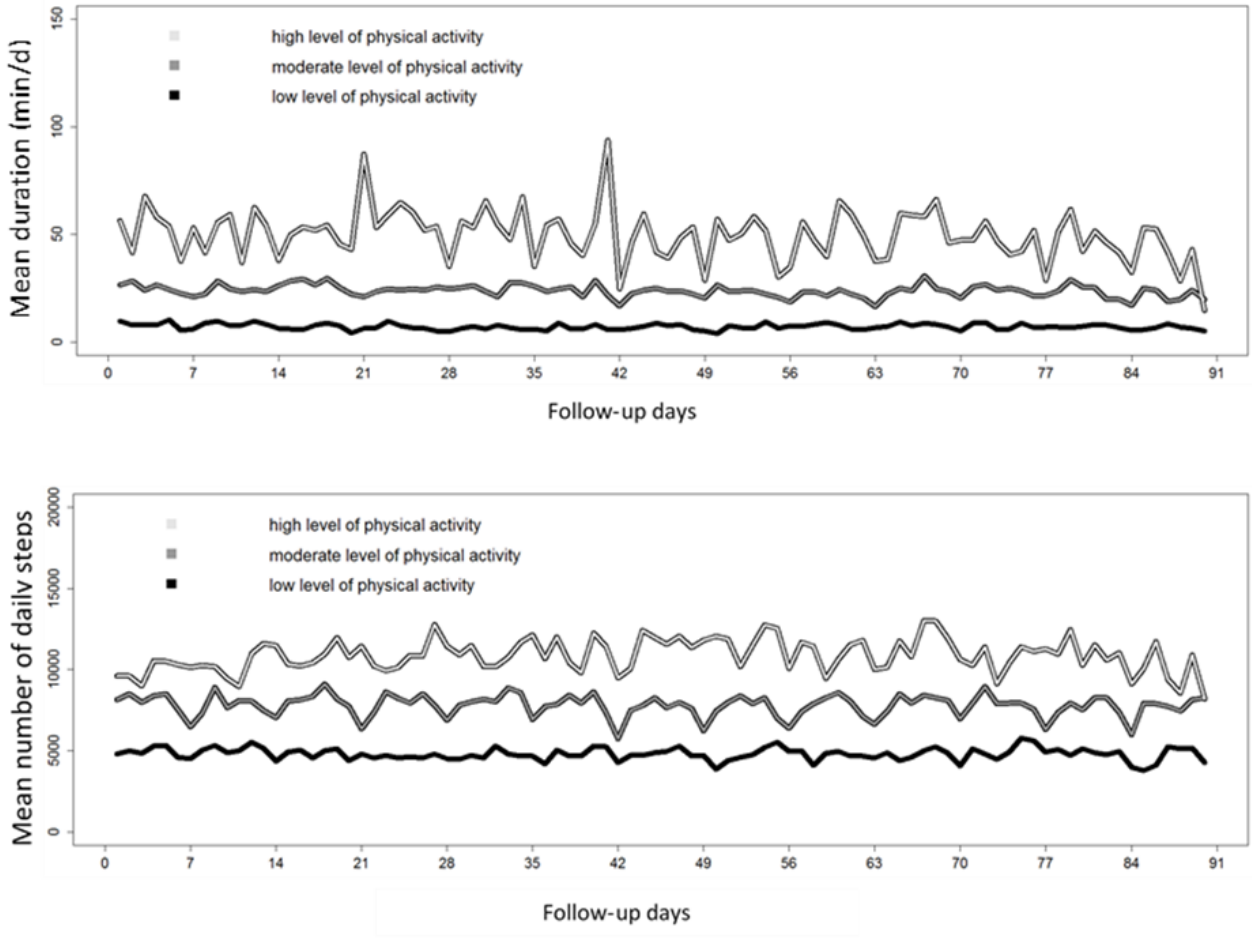


Table 3. Comparison between patients in the low activity cluster and other patients.

Factors associated with a low activity	Clusters according to moderate to vigorous activity duration				Clusters according to number of steps per day			
	Low cluster (N=85)	Other clusters (N=72)	OR[a] (95% CI)	*P* value	Low cluster (N=60)	Other clusters (N=97)	OR (95% CI)	*P* value
Type of arthritis, axSpA[b], n (%)	41 (48)	33 (46)	0.94 (0.79-1.12)	.51	27 (45)	47 (48)	0.97 (0.82-1.15)	.74
Sex, female, n (%)	49 (58)	51 (71)	0.86 (0.72-1.02)	.09	39 (65)	61 (63)	1.00 (0.84-1.18)	.97
Age in years, ≥60 years, n (%)	16 (19)	6 (8)	1.12 (0.84-1.49)	.44	15 (25)	7 (7)	0.97 (1.00-1.69)	.09
BMI[c], kg/m^2, mean (SD[d])	26.1 (5.0)	23.9 (3.5)	1.13 (1.11-1.15)[e]	*.007* [f]	26.5 (5.4)	24.2 (3.5)	1.12 (1.11-1.14)[e]	*.007*
Disease duration, years, mean (SD)	12.3 (9.2)	8.5 (8.2)	1.01(1.00-1.02)	.08	12.1 (9.0)	9.6 (8.8)	1.00 (0.99-1.01)	.79
Employment status, employed, n (%)	63 (74)	57 (79)	1.04 (0.84-1.29)	.74	40 (67)	80 (82)	0.97 (0.78-1.19)	.74
Biological treatment, present, n (%)	54 (64)	34 (47)	1.10 (0.94-1.30)	.23	41 (68)	47 (48)	1.18 (1.01-1.38)	*.04*
PtGA[g] at baseline, mean (SD)	3.6 (2.5)	3.0 (2.2)	1.02 (0.99-1.06)	.15	3.7 (2.5)	3.1 (2.3)	1.02 (0.98-1.05)	.32

[a]OR: odds ratio of low versus moderate to high activity levels.

[b]axSpA: axial spondyloarthritis.

[c]BMI: body mass index.

[d]SD: standard deviation.

[e]OR expressed for an increase in BMI of 5 units,

[f]*P* value ≤.05 are in italics.

[g]PtGA: patient global assessment.

Discussion

Principal Findings

This study brought interesting and original information by exploring longitudinally and objectively physical activity in RA and axSpA using a mobile device. Levels of physical activity were similar in patients with RA and in patients with axSpA, both regarding total and moderate to vigorous physical activity (MVPA), but only infrequently attained the levels recommended by the WHO [7]. A higher BMI was the main factor related to a low level of physical activity. Overall, measuring physical activity using a mobile activity tracker in patients with inflammatory joint disease appeared feasible and acceptable.

Strengths and Weaknesses

This study has strengths and weaknesses. The patients' population may not be representative as indicated by high schooling and job levels and by wide use of biologics treatment. However, this may be an inherent bias when studying mobile devices that necessitate recent and powerful mobile phones [29]. Moreover, the high number of included patients makes this study one of the most ambitious studies using a mobile device in the rheumatology field [34]. The control group was included for indicative purposes only, and therefore, was of small size and did not allow statistical comparisons. However, physical activity in controls was close to the general French population activity [35], which tends to strengthen the validity of the physical activity assessment.

Physical activity may have been misjudged. Although patients were not instructed to perform more physical activity, they get feedback on their activity and may have increased it, because of the motivational aspect of the activity tracker in itself [19,36]. However, the longitudinal analyses did not confirm this bias; we would anticipate that this effect would have decreased over 3 months because of progressive loss of motivation, whereas all trajectories were stable over time. The K-means method used to identify trajectories may nevertheless have underestimated changes [37]. However, no assumptions about trajectories are needed before running the algorithm. Another limitation of the K-means method is to find "spherical" clusters, which have a similar size or variance. In case the population comprised groups with different variances, the K-means method would have difficulty identifying the correct clusters. Analyses may also have been biased by different factors, which may influence physical activity, such as comorbidities, and the type of job. However, patients had very few comorbidities, and only few patients had a manual work. Finally, the alpha risk was not controlled, which may be a limitation of the analysis.

Although activity trackers have been validated in other studies [38,39], estimation of physical activity using a mobile tracker may have some limitations. Nonwalking activities, such as cycling, bodybuilding, and dancing, may have been underestimated. However, this bias is inherent to the measuring method as these activities do not imply a number of steps. Conversely, physical activity may have been overestimated by the tracker (eg, because of arm movement), and some false positive steps may have been recorded [40]. The use of a mobile activity tracker may also be limited by the need of owning a recent mobile phone, having Internet access, and being a minimum familiar with technology.

This study explored different activity patterns. The proportion of morning steps was an original pattern, aiming to reflect the morning stiffness. Unfortunately, this analysis was hampered by the difficulty of defining a precise wake-up time. Morning stiffness may directly influence physical activity by delaying the beginning of daily activities, but in this study, as wake-up time was not recorded, it was difficult to analyze the potential reduction of physical activity just after waking up. The detection of morning stiffness using mobile devices should be further explored.

Which Physical Activity Level to Recommend?

We evidenced that insufficient levels of physical activity were performed both for total physical activity (including slow walking) and MVPA. This indicates that patients with inflammatory joint disease may be both not walking enough and not performing enough aerobic exercises. These findings were confirmed when comparing patients with RA and axSpA with the small sample of healthy persons included. Patients had 20% less total activity duration and daily steps compared with controls. This difference increased to 40% for moderate to vigorous activity duration. However, intense physical activity may not always be increased because of medical issues or aging. It may be more relevant to promote the increase of the total amount of physical activity, including with a low level. Increasing evidence suggests that the practice of all types of physical activity, including low intensity ones, is beneficial to health [41]. Targeting an increase in low to moderate physical activity, such as walking, appears reasonable in patients with rheumatic diseases [42]. These are important data to take into account when discussing physical activity with our patients.

Although the WHO recommendations rest on minutes per week of activity, the message is often transformed into 10,000 steps per day (particularly, when using activity trackers). We show here that these objectives are not equivalent. In this study, 27% of patients (vs 42% of controls) were active enough according to the WHO recommendations (in terms of duration per week of moderate to vigorous activity), but only 10% of patients were above the threshold of 10,000 steps per day. The assessment of physical activity in minutes per week of moderate to vigorous activity may be an achievable and motivating objective for patients with inflammatory joint disease. Activity tracker companies may want to revisit their presentation of physical activity achievements (eg, on their apps). The threshold of both these recommendations may also be too much elevated and difficult to reach, even for healthy persons. Indeed, studies indicate that 58% and 65% of the French and the European general populations, respectively, perform enough physical activity [35,43]. The threshold of 7000 to 8000 steps per day could be more appropriate in elderly or disabled persons [44]. When considering the threshold of 7000 steps per day, 50.3% of the patients would have fulfilled the recommendations. In previous studies, levels of physical activity in patients with inflammatory joint disease were highly variable [11,15,16,45]. Some studies found that more patients were in accordance with the WHO recommendations [11,14,16]. Methods used to assess

the intensity of physical activity varied across studies. This study estimated physical activity from the number of steps per minute, whereas in previous studies, the intensity was mostly estimated from energy expenditure (from questionnaires [46] or multisensor devices [11,45]) or from counts per minute (from conventional accelerometers) [12,16]. Moreover, previous studies were conducted over short durations, and participants may have substantially increased their physical activity during those studies.

Similar levels of physical activity were found in RA and axSpA after adjustment on sex and age. A lower level of physical activity may have been expected in RA patients. Indeed, RA frequently affects the feet, which may directly impact physical activity. These surprising results may be partly explained by the fact that the studied population had a stable disease, was frequently treated with biologics, and had good physical function. Due to its subject, this study may have selected less severe patients. Nevertheless, it is interesting to note that physical activity should be encouraged in both patients with RA and with axSpA.

Risk Factors for a Low Physical Activity

The only factor explaining a low physical activity level in these analyses was a higher BMI. Previous studies found various relations between physical activity and BMI [46,47] and highlighted other factors of inactivity, such as female sex, older age, and lack of motivation [32,48]. Our results suggest that interventions to promote physical activity should target overweight patients. However, given demographic variables did not explain low physical activity, all patients should be encouraged to increase their physical activity. Factors related to a low physical activity level were not explored separately in RA and axSpA subgroups, particularly for disease-specific scores (eg, BASDAI or DAS28). However, physical activity was similar in both subgroups, and the type of arthritis was not related to the low physical activity cluster. The link between physical activity and disease activity should be further studied.

Feasibility and Acceptance of the Use of an Activity Tracker

Finally, this study demonstrated the feasibility of wearing a mobile device over 3 months in patients with inflammatory joint disease. Although patients with inflammatory joint disease may encounter difficulties that are specific to their disease (eg, to handle the device may be difficult because of hand arthritis) [49], patients wore the device 88% of the days, and many were considering pursuing its use. This study indicates that there are no specific barriers linked to RA or axSpA for using activity trackers. Patients were considered adherent enough to be analyzed if at least 60 out 90 days of physical activity were recorded. This cutoff was selected based on a balance between representativeness of physical activity in each patient and representativeness of the study population. However, nonadherence to the activity tracker was not negligible, leading to 11% of patients not being analyzed. This may reflect the constraint of wearing a mobile device every day over a long period of time. Contrary to conventional accelerometers, mobile activity trackers may appear more acceptable to wear over time. Their availability, their ease of use, and their interactive and playful interface make them good devices to measure physical activity in patients [50]. Activity trackers should be further explored as motivational tools to enhance physical activity in inflammatory joint diseases, as indicated in other chronic conditions [19].

Conclusion

Patients with RA and axSpA do not perform enough physical activity. Only 27% of patients met the WHO recommendations, when assessed with an activity tracker. The objective of 150 min per week of MVPA appears more feasible than the objective of 10,000 steps per day. Patients should be encouraged to perform more physical activity in line with the WHO recommendations. Patients with RA and axSpA had similar levels of physical activity both regarding total and moderate/vigorous physical activity, recorded with a mobile activity tracker, despite the differences between these two diseases. This study has shown the feasibility and the interest of mobile devices in rheumatology research; longer-term studies are needed.

Acknowledgments

The authors thank the following physicians who included patients in the study: F Bailly: Rheumatology Department, Pitié Salpêtrière hospital, Paris; S Dadoun: Rheumatology Department, Pitié Salpêtrière hospital, Paris; L Morardet: Rheumatology Department, Pitié Salpêtrière hospital, Paris; S Rozenberg: Rheumatology private practice, Paris; and B Banneville: Rheumatology private practice, Paris. The authors also thank B Granger, Biostatistics Department, Pitié Salpêtrière hospital, Paris, for his help with statistical analyses.

The study was conducted under the umbrella of the French Society of Rheumatology (eHealth section). This study was supported by unrestricted academic grants from Lilly France, Pfizer France and BMS France.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Acceptability regarding use of a connected activity tracker: questionnaire derived from 10 interviews of RA and axSpA patients and results.

[PDF File (Adobe PDF File), 20KB - mhealth_v6i1e1_app1.pdf]

References

1. Braun J, van den Berg R, Baraliakos X, Boehm H, Burgos-Vargas R, Collantes-Estevez E, et al. 2010 update of the ASAS/EULAR recommendations for the management of ankylosing spondylitis. Ann Rheum Dis 2011 Jun;70(6):896-904 [FREE Full text] [doi: 10.1136/ard.2011.151027] [Medline: 21540199]
2. Gaujoux-Viala C, Gossec L, Cantagrel A, Dougados M, Fautrel B, Mariette X, French Society for Rheumatology. Recommendations of the French Society for Rheumatology for managing rheumatoid arthritis. Joint Bone Spine 2014 Jul;81(4):287-297 [FREE Full text] [doi: 10.1016/j.jbspin.2014.05.002] [Medline: 24986683]
3. Sallis JF, Bull F, Guthold R, Heath GW, Inoue S, Kelly P, Lancet Physical Activity Series 2 Executive Committee. Progress in physical activity over the Olympic quadrennium. Lancet 2016 Sep 24;388(10051):1325-1336. [doi: 10.1016/S0140-6736(16)30581-5] [Medline: 27475270]
4. Reis RS, Salvo D, Ogilvie D, Lambert EV, Goenka S, Brownson RC, Lancet Physical Activity Series 2 Executive Committee. Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet 2016 Sep 24;388(10051):1337-1348. [doi: 10.1016/S0140-6736(16)30728-0] [Medline: 27475273]
5. Baillet A, Zeboulon N, Gossec L, Combescure C, Bodin L, Juvin R, et al. Efficacy of cardiorespiratory aerobic exercise in rheumatoid arthritis: meta-analysis of randomized controlled trials. Arthritis Care Res (Hoboken) 2010 Jul;62(7):984-992 [FREE Full text] [doi: 10.1002/acr.20146] [Medline: 20589690]
6. Verhoeven F, Tordi N, Prati C, Demougeot C, Mougin F, Wendling D. Physical activity in patients with rheumatoid arthritis. Joint Bone Spine 2016 May;83(3):265-270. [doi: 10.1016/j.jbspin.2015.10.002] [Medline: 26639220]
7. World Health Organization. Apps.WHO. Geneva, Switzerland: World Health Organization; 2010. Global recommendations on physical activity for health Internet URL: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf [WebCite Cache ID 6rlFx0WIeISBN9789241599979]
8. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
9. Choi BC, Pak AW, Choi JC, Choi EC. Daily step goal of 10,000 steps: a literature review. Clin Invest Med 2007;30(3):E146-E151. [Medline: 17716553]
10. Johnman C, Mackie P, Sim F. 10,000 steps into the digital age. Public Health 2017 Aug;149:A1-A3. [doi: 10.1016/j.puhe.2017.06.018] [Medline: 28709683]
11. Swinnen TW, Scheers T, Lefevre J, Dankaerts W, Westhovens R, de Vlam K. Physical activity assessment in patients with axial spondyloarthritis compared to healthy controls: a technology-based approach. PLoS One 2014;9(2):e85309 [FREE Full text] [doi: 10.1371/journal.pone.0085309] [Medline: 24586239]
12. Prioreschi A, Hodkinson B, Avidon I, Tikly M, McVeigh JA. The clinical utility of accelerometry in patients with rheumatoid arthritis. Rheumatology (Oxford) 2013 Sep;52(9):1721-1727. [doi: 10.1093/rheumatology/ket216] [Medline: 23804220]
13. O'Dwyer T, O'Shea F, Wilson F. Decreased physical activity and cardiorespiratory fitness in adults with ankylosing spondylitis: a cross-sectional controlled study. Rheumatol Int 2015 Nov;35(11):1863-1872. [doi: 10.1007/s00296-015-3339-5] [Medline: 26254884]
14. Haglund E, Bergman S, Petersson IF, Jacobsson LT, Strömbeck B, Bremander A. Differences in physical activity patterns in patients with spondylarthritis. Arthritis Care Res (Hoboken) 2012 Dec;64(12):1886-1894 [FREE Full text] [doi: 10.1002/acr.21780] [Medline: 22745014]
15. Katz P, Margaretten M, Gregorich S, Kaplan S, Rush S, Trupin L. Acrabstracts. 2015. A Randomized Controlled Trial for a Physical Activity Intervention for RA Fatigue abstract URL: http://acrabstracts.org/abstract/a-randomized-controlled-trial-for-a-physical-activity-intervention-for-ra-fatigue/ [accessed 2017-04-19] [WebCite Cache ID 1492636577557666]
16. van GS, Boonen A, van der Heijde D, Heuft L, Luime J, Spoorenberg A, et al. Accelerometer quantification of physical activity and activity patterns in patients with ankylosing spondylitis and population controls. J Rheumatol 2015 Dec;42(12):2369-2375. [doi: 10.3899/jrheum.150015] [Medline: 26523021]
17. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
18. Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]
19. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/sevice (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]

20. US National Library of Medicine. Clinicaltrials.gov. MONITOR-OA: Using Wearable Activity Trackers to Improve Physical Activity in Knee Osteoarthritis URL: https://www.clinicaltrials.gov/ct2/show/NCT02315664?term=Monitor-oa&rank=1 [WebCite Cache ID 6rlGEQQWm]
21. Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R, American College of Rheumatology, European League Against Rheumatism. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: phase 2 methodological report. Arthritis Rheum 2010 Sep;62(9):2582-2591 [FREE Full text] [doi: 10.1002/art.27580] [Medline: 20872596]
22. Rudwaleit M, van der Heijde D, Landewé R, Listing J, Akkoc N, Brandt J, et al. The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection. Ann Rheum Dis 2009 Jun;68(6):777-783. [doi: 10.1136/ard.2009.108233] [Medline: 19297344]
23. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. J Clin Epidemiol 2005 Jun;58(6):595-602. [doi: 10.1016/j.jclinepi.2004.10.018] [Medline: 15878473]
24. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995 Jan;38(1):44-48. [Medline: 7818570]
25. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994 Dec;21(12):2286-2291. [Medline: 7699630]
26. Claudepierre P, Sibilia J, Goupille P, Flipo RM, Wendling D, Eulry F, et al. Evaluation of a French version of the Bath Ankylosing Spondylitis Disease Activity Index in patients with spondyloarthropathy. J Rheumatol 1997 Oct;24(10):1954-1958. [Medline: 9330938]
27. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score with 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S14-S36 [FREE Full text] [doi: 10.1002/acr.20621] [Medline: 22588741]
28. Pincus T, Summey JA, Soraci SA, Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983 Nov;26(11):1346-1353. [Medline: 6639693]
29. Nokia. Health URL: https://health.nokia.com/fr/en/trackers [WebCite Cache ID 6rjmyrXjn]
30. Pinheiro J, Bates D, DebRoy S, Sarkar D. CRAN.R-project. nlme: Linear and Nonlinear Mixed Effects Models URL: http://CRAN.R-project.org/package=nlme [WebCite Cache ID 6rlGIcskp]
31. Genolini C, Falissard B. KmL: a package to cluster longitudinal data. Comput Methods Programs Biomed 2011 Dec;104(3):e112-e121. [doi: 10.1016/j.cmpb.2011.05.008] [Medline: 21708413]
32. Demmelmaier I, Dufour AB, Nordgren B, Opava CH. Trajectories of physical activity over two years in persons with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2016 Aug;68(8):1069-1077. [doi: 10.1002/acr.22799] [Medline: 26784571]
33. Gbif. Vienna, Austria: R Foundation for Statistical Computing; 2015. R: A language and environment for statistical computing URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2017-12-18] [WebCite Cache ID 6vnN87I7D]
34. Espinoza F, Le Blay P, Coulon D, Lieu S, Munro J, Jorgensen C, et al. Handgrip strength measured by a dynamometer connected to a smartphone: a new applied health technology solution for the self-assessment of rheumatoid arthritis disease activity. Rheumatology (Oxford) 2016 May;55(5):897-901. [doi: 10.1093/rheumatology/kew006] [Medline: 26867731]
35. Escalon H, Bossard C, Beck F. inpes.santepubliquefrance.fr. 2008. Baromètre santé nutrition 2008 URL: http://inpes.santepubliquefrance.fr/CFESBases/catalogue/pdf/1270.pdf [accessed 2017-12-18] [WebCite Cache ID 6vnNHWtru]
36. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]
37. Twisk J, Hoekstra T. Classifying developmental trajectories over time should be done with great caution: a comparison between methods. J Clin Epidemiol 2012 Oct;65(10):1078-1087. [doi: 10.1016/j.jclinepi.2012.04.010] [Medline: 22818946]
38. Kooiman T, Dontje M, Sprenger S, Krijnen W, van van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015 Dec;7:24. [doi: 10.1186/s13102-015-0018-5]
39. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
40. O'Connell S, Olaighin G, Quinlan L. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One 2017;12(1):e0169616. [doi: 10.1371/journal.pone.0169616]

41. Matthews CE, Moore SC, Sampson J, Blair A, Xiao Q, Keadle SK, et al. Mortality benefits for replacing sitting time with different physical activities. Med Sci Sports Exerc 2015 Sep;47(9):1833-1840 [FREE Full text] [doi: 10.1249/MSS.0000000000000621] [Medline: 25628179]
42. Mancuso CA, Rincon M, Sayles W, Paget SA. Comparison of energy expenditure from lifestyle physical activities between patients with rheumatoid arthritis and healthy controls. Arthritis Rheum 2007 May 15;57(4):672-678 [FREE Full text] [doi: 10.1002/art.22689] [Medline: 17471544]
43. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
44. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
45. Almeida GJ, Wasko MC, Jeong K, Moore CG, Piva SR. Physical activity measured by the SenseWear Armband in women with rheumatoid arthritis. Phys Ther 2011 Sep;91(9):1367-1376 [FREE Full text] [doi: 10.2522/ptj.20100291] [Medline: 21719635]
46. Hernández-Hernández V, Ferraz-Amaro I, Díaz-González F. Influence of disease activity on the physical activity of rheumatoid arthritis patients. Rheumatology (Oxford) 2014 Apr;53(4):722-731. [doi: 10.1093/rheumatology/ket422] [Medline: 24369410]
47. Stavropoulos-Kalinoglou A, Metsios GS, Smith JP, Panoulas VF, Douglas KM, Jamurtas AZ, et al. What predicts obesity in patients with rheumatoid arthritis? An investigation of the interactions between lifestyle and inflammation. Int J Obes (Lond) 2010 Feb;34(2):295-301. [doi: 10.1038/ijo.2009.220] [Medline: 19859075]
48. Lee J, Dunlop D, Ehrlich-Jones L, Semanik P, Song J, Manheim L, et al. Public health impact of risk factors for physical inactivity in adults with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2012 Apr;64(4):488-493 [FREE Full text] [doi: 10.1002/acr.21582] [Medline: 22278986]
49. Leese J, Tran B, Backman C, Townsend A, Davis A, Jones A, et al. Acrabstracts. 2015. A Qualitative Study of Barriers and Facilitators to Arthritis Patients Use of Physical Activity Monitoring Tools abstract URL: http://acrabstracts.org/abstract/a-qualitative-study-of-barriers-and-facilitators-to-arthritis-patients-use-of-physical-activity-monitoring-tools/ [accessed 2017-07-06] [WebCite Cache ID 6rlGSbRbp]
50. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. JMIR Mhealth Uhealth 2017 Mar 24;5(3):e37 [FREE Full text] [doi: 10.2196/mhealth.6496] [Medline: 28341616]

Abbreviations

ACR/EULAR: American College of Rheumatology / European League Against Rheumatism
anti-CCP: anticyclic citrullinated peptide
ASAS: Assessment of SpondyloArthritis international Society
axSpA: axial spondyloarthritis
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BMI: body mass index
DAS: disease activity score
mHAQ: modified Health Assessment Questionnaire
MRI: magnetic resonance imaging
OR: odds ratio
PtGA: patient global assessment
SD: standard deviation
RA: rheumatoid arthritis
RF: rheumatoid factor
WHO: World Health Organization

Edited by G Eysenbach; submitted 02.05.17; peer-reviewed by E Nikiphorou, T Swinnen, M Vooijs; comments to author 04.07.17; revised version received 15.09.17; accepted 30.10.17; published 02.01.18.

Please cite as:
Jacquemin C, Servy H, Molto A, Sellam J, Foltz V, Gandjbakhch F, Hudry C, Mitrovic S, Fautrel B, Gossec L
Physical Activity Assessment Using an Activity Tracker in Patients with Rheumatoid Arthritis and Axial Spondyloarthritis: Prospective Observational Study
JMIR Mhealth Uhealth 2018;6(1):e1
URL: http://mhealth.jmir.org/2018/1/e1/
doi:10.2196/mhealth.7948
PMID:29295810

Corrigenda and Addenda

Correction: Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy

Yuan Wu[1], MD; Xun Yao[2], MD; Giacomo Vespasiani[3], MD; Antonio Nicolucci[4], MD; Yajie Dong[1], MD; Joey Kwong[5], PhD; Ling Li[5], PhD; Xin Sun[5], PhD; Haoming Tian[1], MM; Sheyu Li[1], MD

[1]Department of Endocrinology and Metabolism, West China Hospital, Sichuan University, Chengdu, China
[2]Department of Academic Affairs, West China School of Medicine, Sichuan University, Chengdu, China
[3]Diabetes Unit, Madonna del Soccorso Hospital, San Benedetto del Tronto (AP), Italy
[4]Center for Outcomes Research and Clinical Epidemiology, Pescara, Italy
[5]Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:
Sheyu Li, MD
Department of Endocrinology and Metabolism
West China Hospital
Sichuan University
37# Guoxue Road
Wuhou District
Chengdu, 610041
China
Phone: 86 13194874843
Fax: 86 2885422982
Email: lisheyu@gmail.com

Related Article:

Correction of: http://mhealth.jmir.org/2017/3/e35/

(JMIR Mhealth Uhealth 2018;6(1):e20) doi:10.2196/mhealth.8789

In "Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy" (JMIR Mhealth Uhealth 2017;5(3):e35), there was an error in Table 2. The "Mean (SD) HbA_{1c}, %: baseline; end; change" for "Rossi 2013" should read "I: 8.4 (NR); 7.9 (NR); –0.5 (NR); C: 8.5 (NR); 8.1 (NR); –0.5 (NR)" instead of "I: 8.4 (0.1); 7.9 (0.1); –0.5 (0.1); C: 8.5 (0.1); 8.1 (0.1); –0.5 (0.1)".

As a result, data were slightly changed as follows:

1. In the Results subsection of the Abstract, the data were changed in 4 places:
   A. "Across 12 included trials involving 974 participants, using app-based interventions was associated with a clinically significant reduction of HbA_{1c} (MD 0.48%, 95% CI 0.19%-0.78%) without excess adverse events";
   B. "Larger HbA_{1c} reductions were noted among patients with type 2 diabetes than those with type 1 diabetes (MD 0.67%, 95% CI 0.30%-1.03% vs MD 0.37%, 95% CI –0.12%-0.86%)";
   C. "Having a complication prevention module in app-based interventions was associated with a greater HbA_{1c} reduction (with complication prevention: MD 1.31%, 95% CI 0.66%-1.96% vs without: MD 0.38%, 95% CI 0.09%-0.67%; intersubgroup P=.01), as was having a structured display (with structured display: MD 0.69%, 95% CI 0.32%-1.06% vs without: MD 0.17%, 95% CI –0.18%-0.53%; intersubgroup P=.05)."
   D. "However, having a clinical decision-making function was not associated with a larger HbA_{1c} reduction (with clinical decision making: MD 0.19%, 95% CI –0.24%-0.63% vs without: MD 0.61%, 95% CI 0.27%-0.95%; intersubgroup P=.14).
2. In the Effects of Mobile App-Based Interventions on HbA_{1c} subsection of the Results, the data were changed in 4 places:
   A. "The use of mobile app-based interventions was associated with a clinically significant HbA_{1c} reduction of 0.48% (95% CI 0.19%-0.78%, I^2=76%, P<.001);

B. "The use of app-based interventions did not achieve statistical significance among patients with T1DM (MD 0.37%, 95% CI –0.12%-0.86%, I^2=86%, P<.001)";
C. Figure 4;
D. Figure 5.

3. In the Effects of Modules, Risks, and Technologies of App-Based Interventions on HbA_{1c} subsection of the Results, data were corrected in the following 5 places:
A. "We noted a greater HbA_{1c} reduction when interventions included a complication prevention module (with complication prevention: MD 1.31%, 95% CI 0.66%-1.96%, I^2=0%, P=.84 vs without: MD 0.38%, 95% CI 0.09%-0.68%, I^2=76%, P<.001; test for subgroup difference P=.01)";
B. "Having a structured display was also associated with a larger HbA_{1c} reduction (with structured display: MD 0.69%, 95% CI 0.32%-1.06%, I^2=63%, P=.008 vs without: MD 0.17%, 95% CI –0.18% to 0.53%, I^2=75%, P=.007; test for subgroup difference P=.05)";
C. "For high-risk interventions with a clinical decision-making function, the reduction of HbA_{1c} was 0.19% (95% CI –0.24%-0.63%, I^2=82%, P=.004), while the reduction was 0.61% (95% CI 0.27%-0.95%, I^2=64%, P=.005) for potential-risk interventions without clinical decision making (test for subgroup difference P=.14)";
D. "Interventions using manual entry showed an associated lower HbA_{1c} reduction without statistical significance (wire connection: MD 0.70%, 95% CI 0.33%-1.07% vs wireless connection: MD 0.53% CI 0.15%-0.92%, I^2 =46%, P=.10 vs manual entry: MD 0.37%, 95% CI –0.12%-0.86%, I^2 =86%, P<.001; test for subgroup difference P=.56)";
E. Figure 6.

4. In the Principal Findings subsection of the Discussion, the data were corrected in 4 places:
A. 1) "The meta-analysis of 12 RCTs demonstrated that app-based interventions were associated with a statistically and clinically significant HbA_{1c} reduction of 0.48% (95% CI 0.19%-0.78%)";
B. 2) "We noted larger HbA_{1c} reductions for patients with T2DM (MD 0.67%, 95% CI 0.30%-1.03%) than those with T1DM (MD 0.37%, 95% CI –0.12%-0.86%)";
C. 3) "The exploratory subgroup analyses showed that having a clinical decision-making function in app-based interventions was not associated with a greater HbA_{1c} reduction (with clinical decision making: MD 0.19%, 95% CI –0.24%-0.63% vs without: MD 0.61%, 95% CI 0.27%-0.95%; intersubgroup P=.14)".

The corrected article will appear in the online version of the paper on the JMIR website on January 15, 2018, together with the publication of this correction notice. Because this was made after submission to PubMed or Pubmed Central and other full-text repositories, the corrected article also has been re-submitted to those repositories.

Please see the corrected data and figures here.

Figure 4. Effects of app-based mobile health interventions on hemoglobin A_{1c} (HbA_{1c}). MD: mean difference.

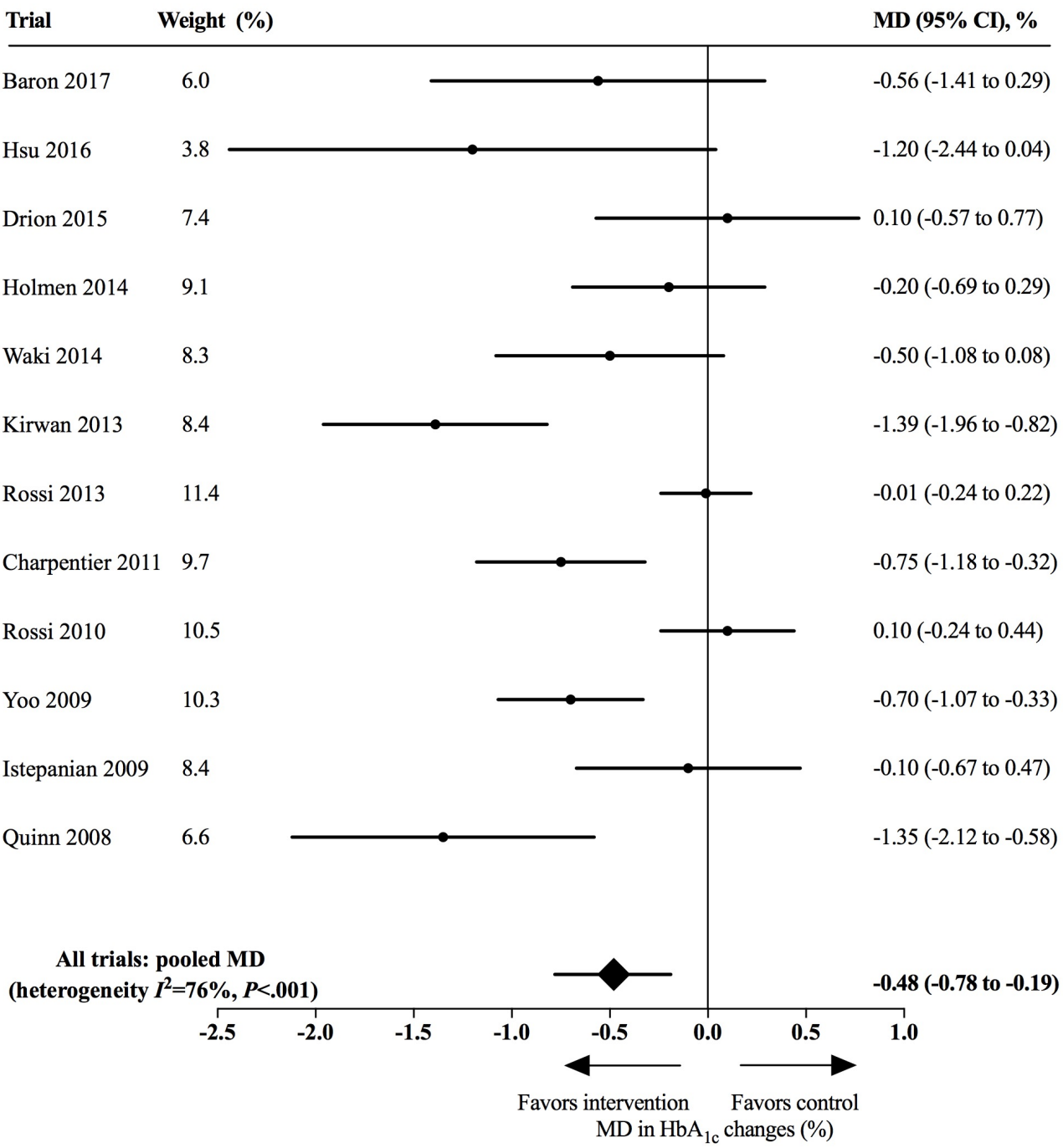


Figure 5. Effects of app-based mobile health interventions on hemoglobin A_{1c} (HbA_{1c}) for patients with type 1 diabetes (T1DM) and type 2 diabetes (T2DM). MD: mean difference.

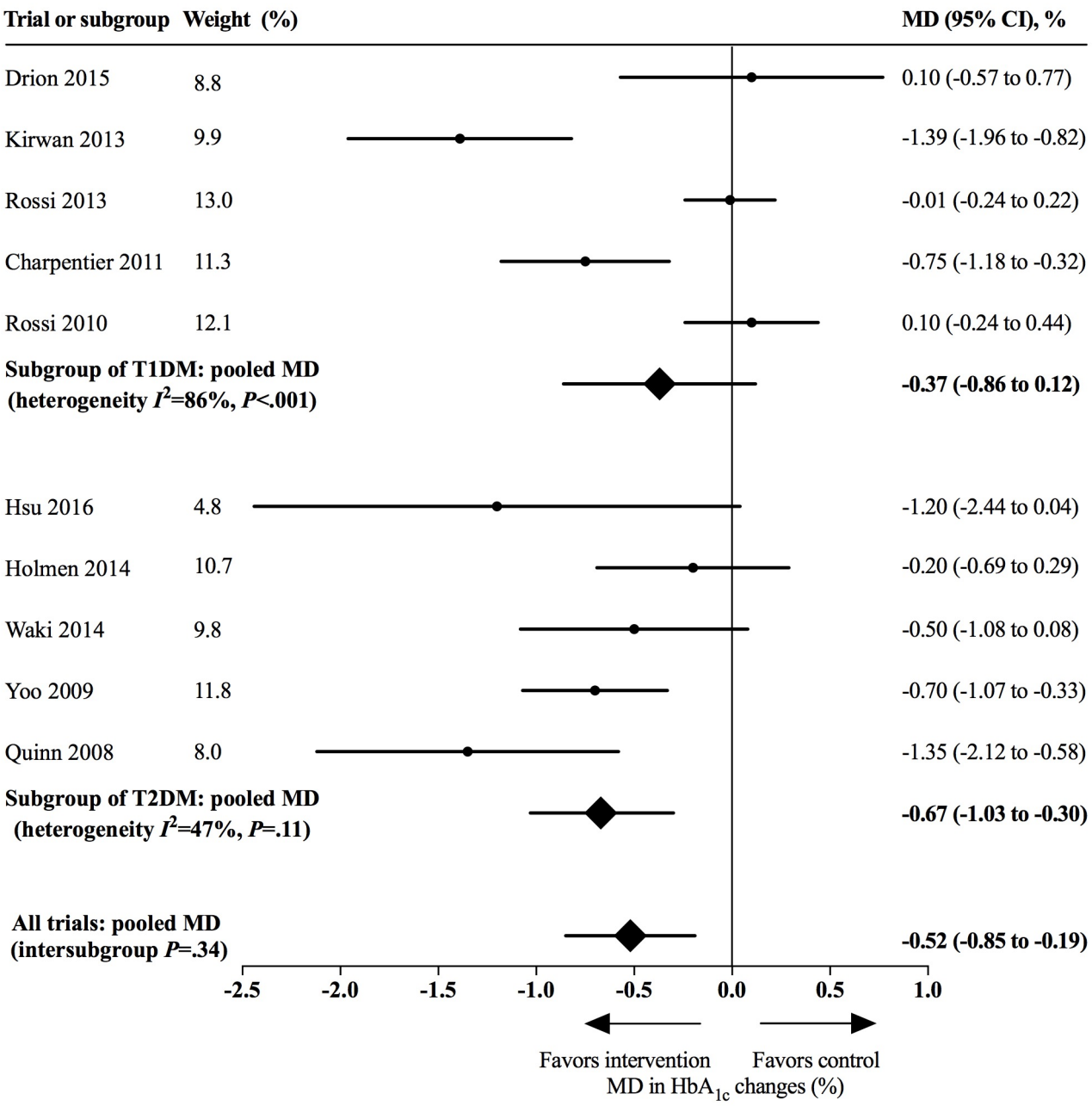
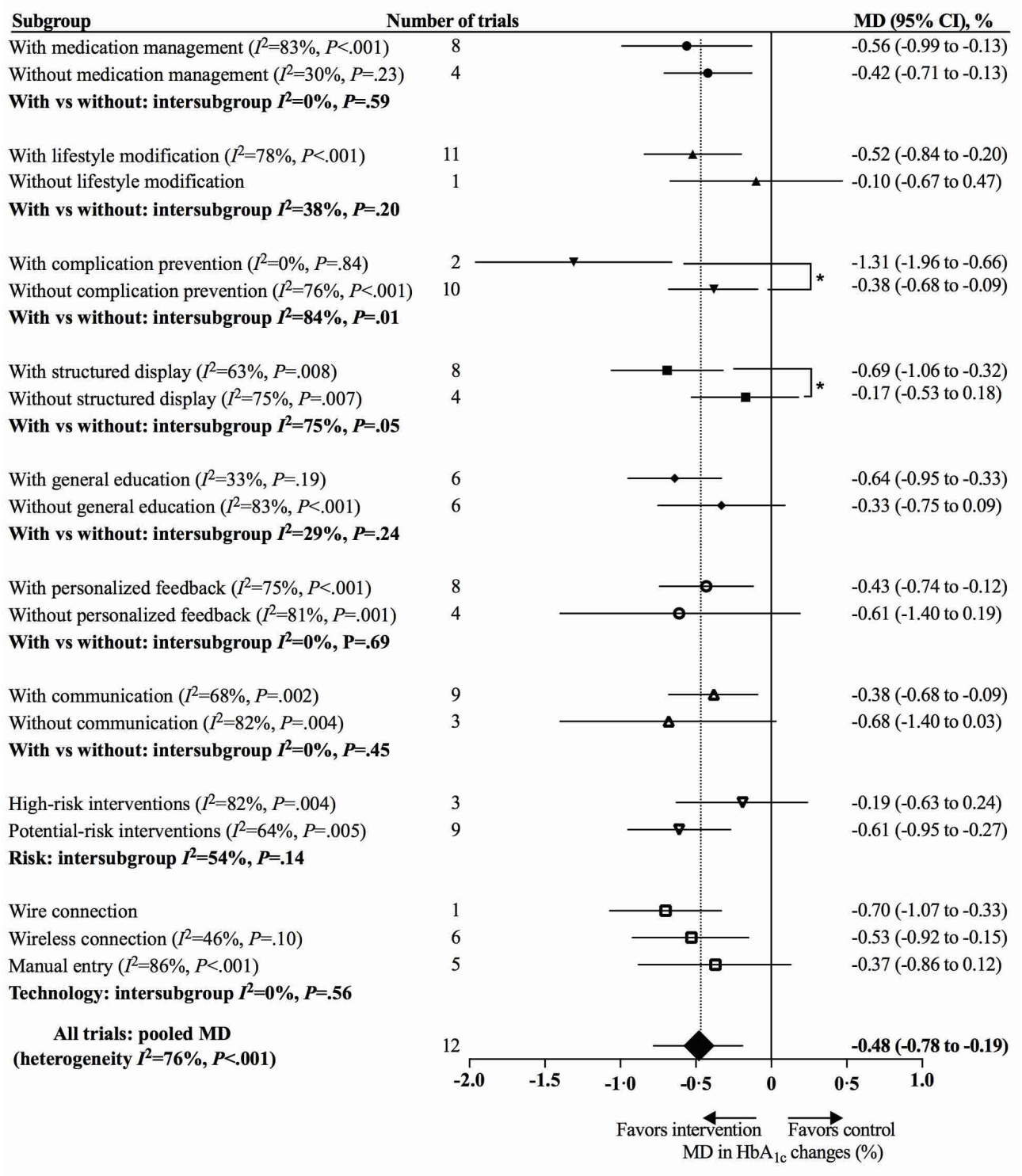


Figure 6. Effects of modules, risks, and technologies of app-based mobile health interventions on hemoglobin A_{1c} (HbA_{1c}). MD: mean difference.



Edited by G Eysenbach; submitted 20.08.17; this is a non–peer-reviewed article;accepted 15.12.17; published 15.01.18.

Please cite as:
Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, Li L, Sun X, Tian H, Li S
Correction: Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy
JMIR Mhealth Uhealth 2018;6(1):e20
URL: http://mhealth.jmir.org/2018/1/e20/
doi:10.2196/mhealth.8789
PMID:29334479

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org