JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 5 (2017), Issue 10 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoints

Original Papers

Viewpoint

A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence

Smisha Agarwal[1,2], MPH, MBA, PhD; Amnesty E Lefevre[1,2], PhD; Alain B Labrique[1,2], MPH, PhD

[1]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[2]Global mHealth Initiative, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
Alain B Labrique, MPH, PhD
Bloomberg School of Public Health
Johns Hopkins University
615 N. Wolfe Street
Baltimore, MD,
United States
Phone: 1 443 287 4744
Fax: 1 410 510 1055
Email: alabriqu@gmail.com

Abstract

Background: Despite the rapid proliferation of health interventions that employ digital tools, the evidence on the effectiveness of such approaches remains insufficient and of variable quality. To address gaps in the comprehensiveness and quality of reporting on the effectiveness of digital programs, the mHealth Technical Evidence Review Group (mTERG), convened by the World Health Organization, proposed the mHealth Evidence Reporting and Assessment (mERA) checklist to address existing gaps in the comprehensiveness and quality of reporting on the effectiveness of digital health programs.

Objective: We present an overview of the mERA checklist and encourage researchers working in the digital health space to use the mERA checklist for reporting their research.

Methods: The development of the mERA checklist consisted of convening an expert group to recommend an appropriate approach, convening a global expert review panel for checklist development, and pilot-testing the checklist.

Results: The mERA checklist consists of 16 core mHealth items that define what the mHealth intervention is (content), where it is being implemented (context), and how it was implemented (technical features). Additionally, a 29-item methodology checklist guides authors on reporting critical aspects of the research methodology employed in the study. We recommend that the core mERA checklist is used in conjunction with an appropriate study-design specific checklist.

Conclusions: The mERA checklist aims to assist authors in reporting on digital health research, guide reviewers and policymakers in synthesizing evidence, and guide journal editors in assessing the completeness in reporting on digital health studies. An increase in transparent and rigorous reporting can help identify gaps in the conduct of research and understand the effects of digital health interventions as a field of inquiry.

(JMIR Mhealth Uhealth 2017;5(10):e136) doi:10.2196/mhealth.6640

KEYWORDS

mHealth; checklist; reporting; digital health; publishing guidelines

Introduction

Over the last decade, there has been a dramatic increase in health programs employing digital tools, such as mobile phones and tablets, to stimulate demand for or the delivery of health care services. This is especially true in low- and middle-income countries, where public health practitioners are tapping into the unprecedented growth in the use of mobile phones to overcome information and communications challenges [1,2]. Donors have rallied around digital approaches, and much has been invested into developing, testing, and deploying digital systems. However, after nearly a decade of concerted efforts, widely available evidence in support of digital health is limited [1,3,4]. As an emergent field, there is substantial variability in the reporting of digital program implementations, evaluations, and outcomes. Inconsistency in reporting is problematic as it limits

policy makers' ability to understand precise program details and extract, compare, and synthesize linkages (if any) between the digital investments and consequent health effects.

To address gaps in the comprehensiveness and quality of reporting on the effectiveness of digital programs, the mHealth Technical Evidence Review Group (mTERG)—an expert committee convened by the World Health Organization (WHO) to advise on approaches to strengthening digital health evidence—proposed guidelines for reporting evidence on the development and evaluation of digital health interventions. These guidelines—presented as the mHealth Evidence Reporting and Assessment (mERA) checklists—were published in March 2016 [5] and have since been widely accessed [1,6-10].

Methods

The design of the mERA checklist followed a systematic process for the development of reporting guidelines [11]. In October 2012, WHO convened an expert working group led by the Johns Hopkins Global mHealth Initiative to develop an approach for the mERA guideline. In December 2012, this working group presented an initial draft of the checklist to a global panel of 18 experts convened by WHO during a 3-day meeting in Montreaux, Switzerland. At this meeting, the approach and checklist underwent intensive analysis for improvement, and a quality of information (QoI) taskforce was established to pilot-test the checklist. After testing by the QoI taskforce, the checklist and associated item descriptions were applied to 10 English language reports to test the applicability of each criterion to a range of existing mHealth literature. Readers may refer to further details about the methodology in the complete manuscript [5].

Results

The mERA checklists comprises 2 components. The core mHealth checklist (see Table 1) identifies a minimum amount of information needed to define what the mHealth intervention is (content), where it is being implemented (context), and how it was implemented (technical features). This checklist may be valuable to researchers in reporting on the program and research results in peer-reviewed journals and reports, to policy makers in consolidating evidence and understanding the quality of information that has been used to generate the evidence, and to program implementers thinking through and selecting core elements for new digital health projects. L'Engle et al [12] applied the mERA checklist to evaluate the quality of evidence on the use of digital health approaches to improving sexual and reproductive health outcomes for adolescents. The study found that, on average, 7 out of 16 (41%) of the core mHealth checklist items were reported on, suggesting a lack of the availability of a clear description of the digital health intervention [12]. During the development and testing phase, the mERA checklist was applied to literature on the use of digital devices in reducing drug stockouts and the use of digital protocols to improve provider adherence to treatment protocols. Interested authors should refer to the definitions and examples for the core mHealth checklist available freely online [5].

Table 1. mHealth Evidence Reporting and Assessment (mERA) core checklist items.

Number	Item
1	Infrastructure
2	Technology platform
3	Interoperability/health information systems (HIS) context
4	Intervention delivery
5	Intervention content
6	Usability/content testing
7	User feedback
8	Access of individual participants
9	Cost assessment
10	Adoption inputs/program entry
11	Limitations for delivery at scale
12	Contextual adaptability
13	Replicability
14	Data security
15	Compliance with national guidelines or regulatory statutes
16	Fidelity of the intervention

Textbox 1. mHealth Evidence Reporting and Assessment (mERA) methodology.

Introduction

1. Rationale/scientific rationale
2. Objectives/hypotheses
3. Logic model/theoretical framework

Methods

4. Study design
5. Outcomes
6. Data collection methods
7. Participant eligibility
8. Participant recruitment
9. Bias
10. Sampling
11. Setting and location
12. Comparator
13. Data sources

Result

14. Enrollment
15. Description of study population
16. Reporting on outcomes

Discussion

17. Summary of evidence
18. Limitations
19. Generalizability
20. Conclusions

Conflicts

21. Funding
22. Ethical considerations
23. Competing interests

Additional criteria for quantitative study methods

24. Confounding
25. Statistical methods
26. Missing data

Additional criteria for qualitative study methods

27. Analytic methods
28. Data validation
29. Reflexivity of account

The methodology checklist (see Textbox 1) outlines 29 items that highlight the key study design features that should be reported by researchers and evaluators of digital health interventions. Authors interested in using this checklist should note that there are other recommended checklists specific to different study designs—for example, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) for observational studies [13] and Consolidated Standards of Reporting Trials (CONSORT) for randomized trials [14]. We recommend that the core mHealth checklist be used in conjunction with these extant checklists based on the appropriate research study design that is being reported. However, we also recognize that a number of digital health studies that are being conducted to evaluate early-stage digital health interventions are more exploratory in nature, and the extant guidelines might not be as relevant to them. In such cases, the authors may decide

to use the mERA methodology checklist, developed to be study-design agnostic, for reporting on the study design and results. A detailed explanation of the mERA methodology checklist items is available as a Web appendix [5].

Discussion

We present an overview of the mERA checklist. For details about each of the checklist items under the core checklist items and the methodology items, we refer the readers to the complete publication [5]. The mERA checklist marks the culmination of several years of multiinstitutional collaborations, led by WHO, to determine appropriate standards for reporting on digital health evidence—standards that not only address issues of methodological and reporting rigor but also are responsive to the current state of the digital health space. We recognize that the digital health space is constantly evolving and is somewhat unique in its multidisciplinary nature, borrowing approaches from the fields of health care and technology and often engaging innovators who are unfamiliar with scientific methodologies. The mERA core and methodology checklists were pragmatically developed to be useful to a wide audience of innovators. We expect that the detailed explanations and examples make the checklist easy to use for individuals with varying levels of experience in academic reporting.

Even as the numbers of digital health interventions continue to increase, the evidence to support such interventions remains sparse. Without the support and shared commitment of the diverse digital health community in advancing the quality of evidence, the state of the much-critiqued “pilotitis” in mHealth will not change [15]. Transparency in the reporting of what constitutes a digital health intervention and clarity on evaluation methods are both critical to determining whether the digital strategy might be scalable to an entire population. In order to support the widespread adoption of the checklist, we encourage digital health researchers and program managers to ensure conformity with the checklist items. Additionally, we would like to call upon editors of journals publishing mHealth literature to encourage the use of the mERA checklist by presenting the link to the guidelines under Instructions to Authors and inclusion of a statement in the manuscript that “this manuscript was developed in conformity with the recommended criteria for reporting digital health as described in the mERA guidelines.”

Conflicts of Interest

None declared

References

1. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
2. 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]
3. 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]
4. 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]
5. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
6. Altmetric. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist: overview of attention for article published in British Medical Journal. URL: https://www.altmetric.com/details/6208109 [accessed 2017-02-28] [WebCite Cache ID 6ochPcPLv]
7. Fottrell E, Jennings H, Kuddus A, Ahmed N, Morrison J, Akter K, et al. The effect of community groups and mobile phone messages on the prevention and control of diabetes in rural Bangladesh: study protocol for a three-arm cluster randomised controlled trial. Trials 2016 Dec 19;17(1):600 [FREE Full text] [doi: 10.1186/s13063-016-1738-x] [Medline: 27993166]
8. Rothstein JD, Jennings L, Moorthy A, Yang F, Gee L, Romano K, et al. Qualitative assessment of the feasibility, usability, and acceptability of a mobile client data app for community-based maternal, neonatal, and child care in rural Ghana. Int J Telemed Appl 2016;2016:2515420 [FREE Full text] [doi: 10.1155/2016/2515420] [Medline: 28070186]
9. Pagliari C. Digital support for childbirth in developing countries: seeds of hope in an evidential desert. JAMA Pediatr 2016 Aug 01;170(8):737-739. [doi: 10.1001/jamapediatrics.2016.1010] [Medline: 27322894]
10. Torous J, Firth J, Mueller N, Onnela JP, Baker JT. Methodology and reporting of mobile heath and smartphone application studies for schizophrenia. Harv Rev Psychiatry 2017 Feb 23. [doi: 10.1097/HRP.0000000000000133] [Medline: 28234658]
11. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010 Feb;7(2):e1000217 [FREE Full text] [doi: 10.1371/journal.pmed.1000217] [Medline: 20169112]
12. L'Engle KL, Mangone ER, Parcesepe AM, Agarwal S, Ippoliti NB. Mobile phone interventions for adolescent sexual and reproductive health: a systematic review. Pediatrics 2016 Sep;138(3). [doi: 10.1542/peds.2016-0884] [Medline: 27553221]

13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
14. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
15. Labrique A, Vasudevan L, Chang LW, Mehl G. H_pe for mHealth: more “y” or “o” on the horizon? Int J Med Inform 2013 May;82(5):467-469 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.016] [Medline: 23279850]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
mERA: mHealth Evidence Reporting and Assessment
mTERG: mHealth Technical Evidence Review Group
QoI: quality of information
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
WHO: World Health Organization

Edited by CL Parra-Calderón; submitted 10.09.16; peer-reviewed by K Reuter, P Seth, C Auschra, S Clough, A Garg, R Scott, M Honary, I Montagni; comments to author 23.10.16; revised version received 10.11.16; accepted 28.03.17; published 06.10.17.

Please cite as:
Agarwal S, Lefevre AE, Labrique AB
A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence
JMIR Mhealth Uhealth 2017;5(10):e136
URL: https://mhealth.jmir.org/2017/10/e136/
doi:10.2196/mhealth.6640
PMID:28986340

Viewpoint

Mobile Health Apps in OB-GYN-Embedded Psychiatric Care: Commentary

Aydan Mehralizade[1], MPH; Shayna Schor[1]; Chad M Coleman[1], MPH; Claire E Oppenheim[1], MPH; Christy A Denckla[2], PhD; Christina PC Borba[3], MPH, PhD; David C Henderson[3], MD; James Wolff[4], MD, MPH; Sarah Crane[1,3], MD; Pamela Nettles-Gomez[1,3], MHA; Avik Pal[5]; Snezana Milanovic[1], MD, MSc

[1]Boston Medical Center, Boston, MA, United States

[2]Harvard TH Chan School of Public Health, Harvard University, Cambridge, MA, United States

[3]Boston University School of Medicine, Boston, MA, United States

[4]Boston University School of Public Health, Boston, MA, United States

[5]CliniOps, Fremont, CA, United States

Corresponding Author:
Snezana Milanovic, MD, MSc
Boston Medical Center
1 Boston Medical Center Pl, Boston, MA 02118
Boston, MA,
United States
Phone: 1 617 414 1917
Fax: 1 617 414 1910
Email: snezana.milanovic@bmc.org

Abstract

This paper explores the potential benefits of the use of mobile health (mHealth) apps in obstetrician-gynecologist (OB-GYN)-embedded psychiatric clinics in the United States. First, we highlight the increasing trend of integrating mental health care within the OB-GYN context. Second, we provide examples of successful uses of mHealth in the global health context and highlight the dearth of available research in the United States. Finally, we provide a summary of the shortcomings of currently available apps and describe the upcoming trial of a novel app currently underway at the Mother-Child Wellness Clinical and Research Center at Boston Medical Center.

(JMIR Mhealth Uhealth 2017;5(10):e152) doi:10.2196/mhealth.7988

KEYWORDS

mHealth; eHealth; embedded psychiatric clinic; postpartum depression; mental health; OB-GYN; global health; reproductive health

Mental Health Care in the OB-GYN Setting

Overview

Women are particularly vulnerable to increased mental health problems in the peripartum and antepartum periods, and again at menopause [1]. Unfortunately, these symptoms of mental illness go largely undiagnosed and untreated because of time constraints and lack of trained providers [2,3]. Obstetrician-gynecologists (OB-GYNs) are often the only regular health care providers for women. OB-GYNs see a full third of all nonillness-related (ie, prophylactic) visits for women under 65 [4]. Among women with mental illness, the highest rates of depression occur in women of lower socioeconomic status, minority women, and immigrant women. These women tend to seek primary care within OB-GYN settings [5].

For the abovementioned reasons, health care professionals have been advocating for an integrated care approach that incorporates psychiatric care into the OB-GYN setting [6,7]. Available evidence from randomized trials suggests that integrated care has a significant positive impact on depression outcomes, especially among women of lower socioeconomic status [8]. In addition, researchers have estimated that mental health care integrated into the primary care setting can save up to US $48 billion annually in the United States [9].

mHealth Care in the Global OB-GYN Context

A key emerging technology that could potentially advance integrated care is mobile health (mHealth), particularly mHealth

apps. mHealth is defined as "medical and public health practices supported by mobile devices including mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [10]. mHealth sits at the intersection of electronic health and mobile phone technology [11]. Mobile phones are the most popular mobile technology used today and their use will only continue to increase. According to estimates by the mobile communications giant Ericsson, by 2019, there will be 9.3 billion mobile subscriptions in the world; 5.6 billion of these will be mobile phone subscriptions [12]. Mobile phones have a clear advantage over many other technologies due to their portability, Internet connectivity, and power to run several apps at once. Tablets connected to the Internet by Wi-Fi, such as the Apple iPad and Google Nexus, are increasingly used in the health field to deliver educational material for patients and medical professionals, run medical algorithms, and provide access to patient data [13].

A mobile app is software that runs on a mobile device [14]. Between 2005 and 2011, the number of mHealth apps launched increased by 30% [15]. As of 2015, there were over 40,000 medical apps available for tablets and mobile phones and over 247 million people had downloaded an mHealth app [16]. mHealth apps provide appointment tracking, mood/behavior tracking, community engagement, patient education, and support for behavior change [17].

A recent meta-analysis found that mHealth apps in the OB-GYN setting have been shown to improve women's health outcomes in the global health context (eg, increased antenatal care attendance and improved breastfeeding practices) [18]. For example, the Mobile Alliance for Maternal Action (MAMA) project provides information to pregnant women based on the woman's estimated date of delivery. MAMA was a 3-year, US $10 million investment cofunded by the United States Agency for International Development, the mHealth Alliance, and the United Nations Foundation, among others, with a goal of delivering health messaging via mobile phones to expectant mothers in Bangladesh, India, and South Africa. MAMA successfully achieved its goals of promoting institutional deliveries, antenatal and postnatal care visit uptake, and exclusive breastfeeding in the three original locations. As evidence of the project's ability to scale up, in 2014, the global coordination office of MAMA closed, and in-country partners took full ownership of their respective projects. Currently, the MAMA model is being used in 54 countries by 161 organizations [19]. In a relatively new field of mHealth, where successful pilot studies struggle to maintain funding, MAMA is a leading example of an effective and sustainable intervention.

The global health literature contains many more examples of successful use of mHealth in women's primary health care settings. In Oro State, Nigeria, a successful mHealth project improved antenatal care appointment adherence and increased diagnosis and treatment of pregnancy complications. This was achieved by giving pregnant women mobile phones and sending them weekly short message service (SMS) text messages with information on potential complications and other educational materials [20]. In Rwanda and Zanzibar, mHealth is reported to have improved maternal health outcomes and reduced neonatal mortality [21,22].

mHealth Care in the OB-GYN Setting in the United States

Despite evidence that the use of mHealth apps in global OB-GYN settings results in improved outcomes, in the United States there is scant evidence on the use of mHealth to improve health outcomes [23]. A study in 2012 found that 59% of patients in emerging markets use at least one mHealth app compared to 35% of patients in the developed world [14]. This gap represents an opportunity for integrated psychiatric services within OB-GYN settings to incorporate mHealth technology, capitalize on the experience gained in the global health field in the past 10 years, and improve mental and reproductive health outcomes of their patients.

The Mother-Child Wellness Clinical and Research Center

Overview

The Mother-Child Wellness Clinical and Research Center (MCWCRC) at Boston Medical Center, founded in 2016, is pursuing this strategy to enhance care for women and children by integrating psychiatric care in the OB-GYN setting. The patient population of Boston Medical Center is diverse in terms of socioeconomic status and ethnic and racial background: 71.52% are nonwhite, 33.3% are Hispanic, and 82.49% are Medicaid patients (detailed demographics of the MCWCRC patient population are provided in Figures 1 and 2). Women who fit into these demographic categories are more likely to report institutional and stigma-related barriers to accessing mental health care when visiting an OB-GYN clinic, despite high interest in receiving some form of psychiatric care [24]. In fact, the no-show rate at the MCWCRC is approximately 50% and patients identify lack of transportation options as a major barrier to accessing care (unpublished internal research).

In order to improve the quality of services offered to its target population, the MCWCRC plans to design and integrate an mHealth app into its continuum of care. The proposed mHealth app will be used to conduct depression screening, reduce loss to follow-up, integrate patients' mental health records with the clinic's existing electronic medical records system (ie, Epic), help patients track their appointments, record behavioral symptoms, provide educational information, engage with the community, and alert providers to patients who may need a mental health intervention.

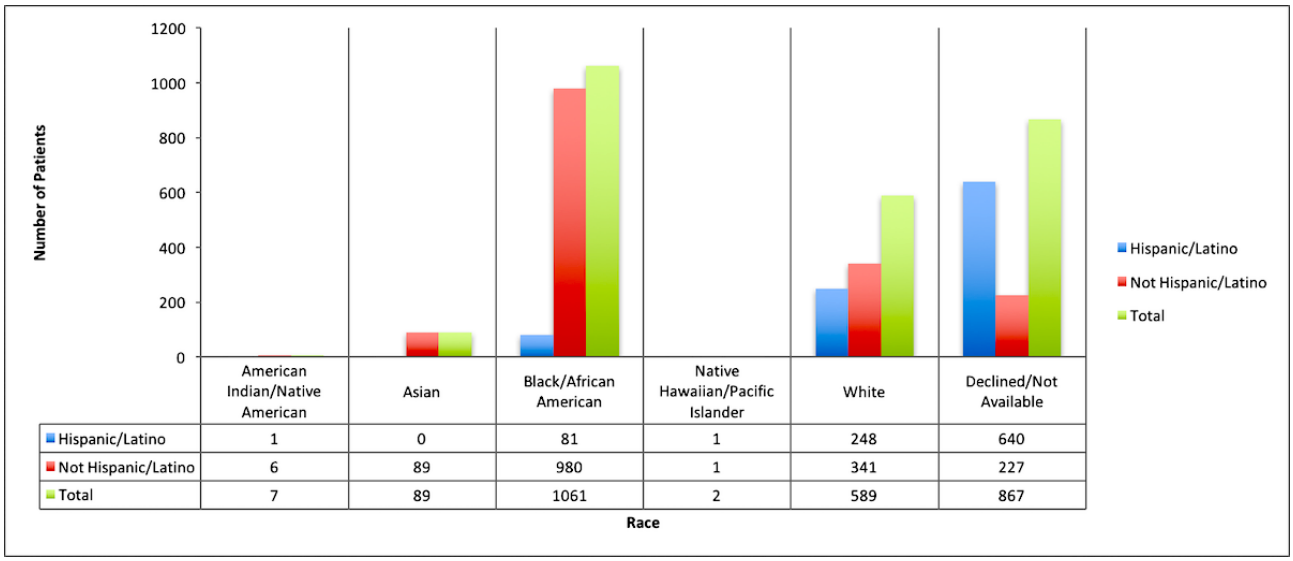
The MCWCRC at Boston Medical Center decided to build a customized mHealth app because the mHealth apps available on the US market do not meet all the needs for an integrated psychiatric facility embedded within an OB-GYN clinic. Apps that track mood and help patients cope with their psychiatric conditions do not provide other essential functions, such as appointment tracking, community engagement, or prenatal and postnatal education. On the other hand, apps aimed specifically at prenatal and postnatal education, appointment tracking, and community engagement lack the crucial aspect of mental health engagement. A brief summary of various apps available on the market is provided in Table 1 below. No information on the

number of subscribers to each app can be found in the published literature.

mHealth Trial at the Mother-Child Wellness Clinical and Research Center

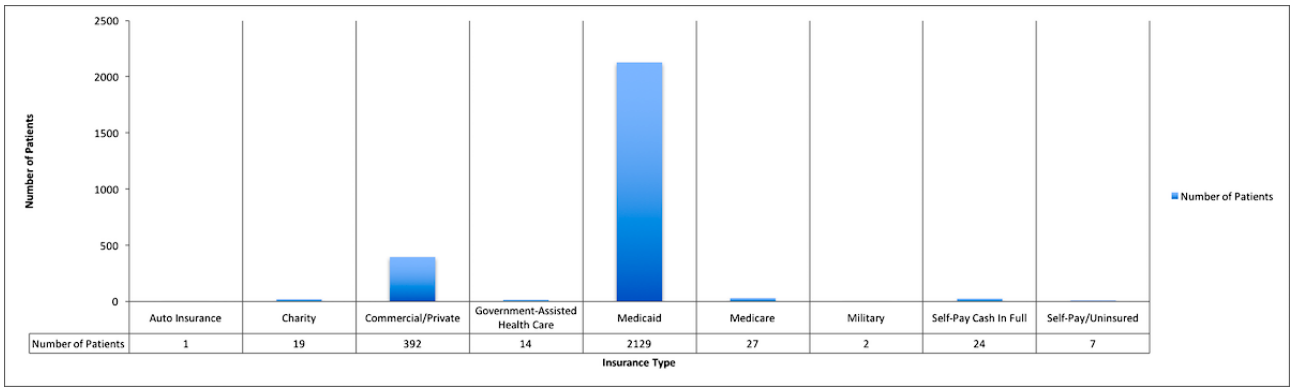
The MCWCRC team, working with CliniOps, a vendor specializing in e-solutions for monitoring and patient management, is implementing a custom-built tablet and mobile phone app at Boston Medical Center [25]. CliniOps has already developed and implemented a mental health-focused app in the global health context and is currently in the process of customizing the app in accordance with the specific needs of the MCWCRC population. The prospective cohort study will take place in two phases: (1) a limited 3-month trial (ie, pilot phase or Phase I), slated to begin in the fall of 2017, and (2) a subsequent clinic-wide trial (ie, Phase II) scheduled to last 12 months.

Figure 1. Race distribution of the 2615 women who delivered at Boston Medical Center between May 1, 2015 and May 1, 2016.



	American Indian/Native American	Asian	Black/African American	Native Hawaiian/Pacific Islander	White	Declined/Not Available
Hispanic/Latino	1	0	81	1	248	640
Not Hispanic/Latino	6	89	980	1	341	227
Total	7	89	1061	2	589	867

Figure 2. Health insurance distribution of the 2615 women who delivered at Boston Medical Center between May 1, 2015, and May 1, 2016.



	Auto Insurance	Charity	Commercial/Private	Government-Assisted Health Care	Medicaid	Medicare	Military	Self-Pay Cash In Full	Self-Pay/Uninsured
Number of Patients	1	19	392	14	2129	27	2	24	7

Table 1. Comparison of prenatal and pregnancy apps available on the market in terms of various functions.

App	US or global	Appointment tracking	Patient education	Community engagement	Behavior/mood tracking	Payment
BabyCenter	Global[a]	X	X	X		In-app purchases
Glow Nurture	Global[a]	X	X	X		In-app purchases
Text4baby	US	X	X			Free
myStrength	US				X	Free
PPD ACT[b]	US				X	Free
MAMA[c]	Global[d]	X	X			Free

[a]Apps that we mostly used in developed English-speaking countries (ie, United States, United Kingdom, Australia, and New Zealand).

[b]PPD ACT: Postpartum Depression: Action Towards Causes and Treatment.

[c]MAMA: Mobile Alliance for Maternal Action.

[d]Apps used in developing countries.

In the pilot phase (Phase I), all pregnant and postpartum patients entering the clinic will be asked to complete an iPad-administered questionnaire, available in both English and Spanish. We anticipate at least 1000 patients will be enrolled in the pilot phase of the study. The iPad questionnaire will collect data on demographics and general psychiatric symptoms. Namely, it will use the following psychiatric scales: Edinburgh Postnatal Depression Scale (EPDS), the Life Events Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (LEC-5) Trauma Scale, and the Andrew Cherry-Oklahoma Co-Occurring Disorders (AC-OK COD) Screen [26-28]. The EPDS is a commonly used, self-screening tool used to identify depression. The LEC-5 Trauma Scale is a self-screening measure meant to identify traumatic events in the respondent's life. The AC-OK COD Screen is a self-administered tool to establish whether the respondent experiences co-occurring mental illness and substance abuse. The contents of the questionnaire and the scales will be read aloud to patients with low literacy levels. If a patient exceeds the EPDS threshold for depression (>10), an automatic educational module will be triggered regarding depression in pregnancy and postpartum. In addition, a care choice menu will be activated asking the patient whether she would like to contact her medical doctor or a case manager, or if she would like to access individual or group therapy sessions. Subsequently, patients will be offered the option of consulting a mental health professional on-site. Furthermore, patients will be asked for their consent to participate in a future research study, as well as whether they would be receptive to downloading a mobile phone app for that purpose. The mobile phone app by CliniOps will provide educational information on gestation and mental health, as well as opportunities for community engagement, appointment reminders, and mood/behavior tracking. After the conclusion of the pilot study, the data will be analyzed, and the respondents' acceptance levels of the mobile phone app will be measured.

During the clinic-wide trial (Phase II), patients will choose between downloading the mobile phone app or receiving follow-up care as usual. The clinic-wide phase will thus be an open-label study; the differences in treatment adherence (ie, reflected in the no-show rate), clinical outcomes (ie, time to depression remission), and health care utilization between the two groups (ie, app users and standard care users) will be measured.

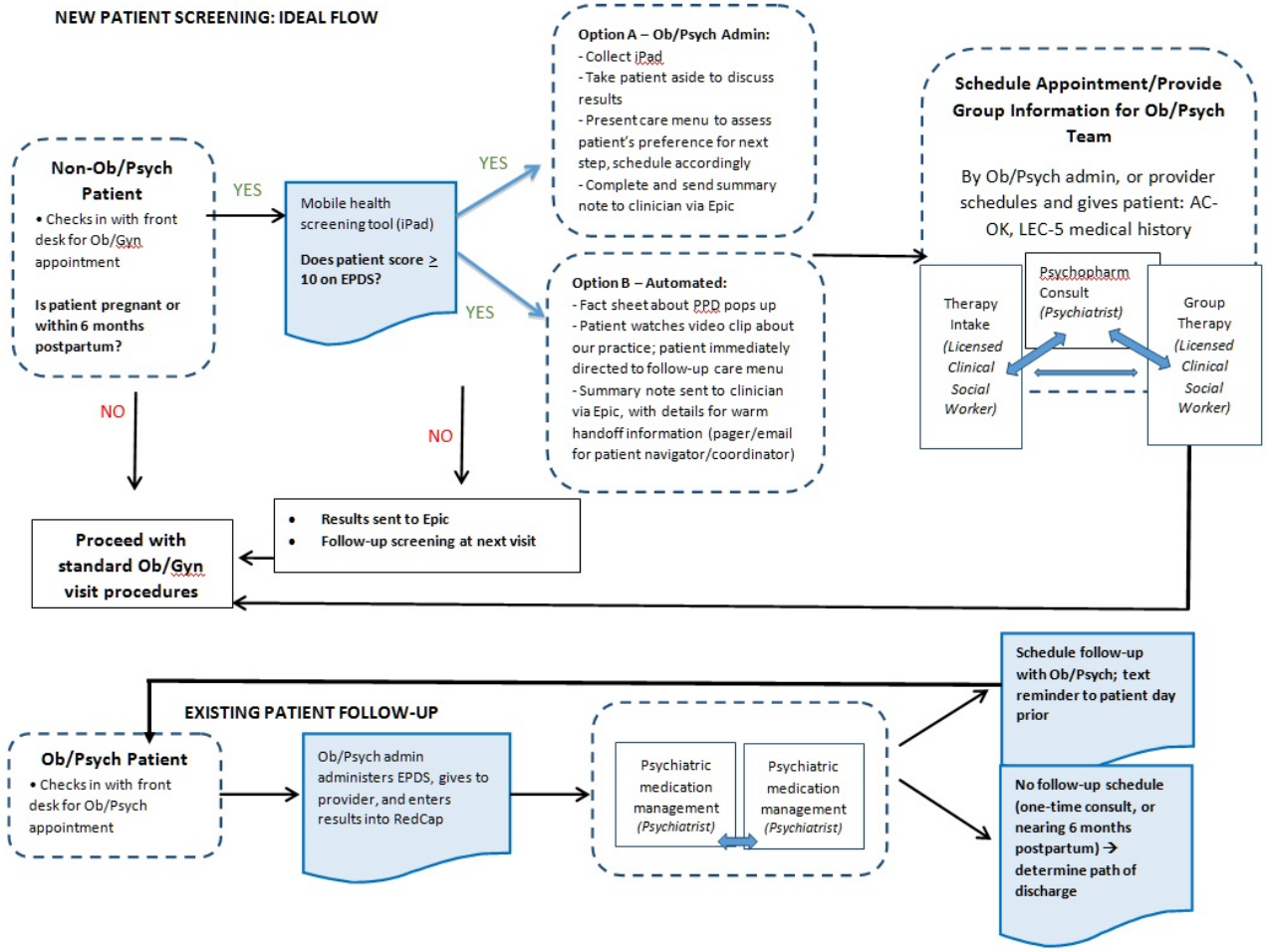
The long-term goal of this mHealth trial (Phase II) is to incorporate the patients' questionnaire data into their Epic record and send screening summary alerts to their respective obstetricians and mental health providers. Figure 3 provides a visual representation of the patient flow during the pilot phase of the trial.

The use of mHealth apps among patient populations should be carefully considered in the context of the socioeconomic setting. This concern is due to the expenses associated with maintaining sufficient data to receive SMS text messages and read educational materials online. In the United States, 68% of adults are estimated to have a mobile phone, while 45% have a tablet [29]. The socioeconomic composition of the patient population must be taken into account when designing a mobile health intervention to ensure equity and reliability of the trial. In addition, patients' educational levels must also be considered. The MCWCRC team in conjunction with CliniOps is cognizant of the literacy level concerns and is committed to making the tablet- and mobile phone-based app accessible to the Boston Medical Center patient population. There is encouraging evidence on this topic in the published literature: people with low self-reported health literacy are likely to seek information from mobile phone-based health apps [30].

In addition, global health experience shows that the uptake of successful pilot studies is often limited [31]. With some rare exceptions, such as the MAMA project, mHealth interventions often fail to scale up and demonstrate long-term utility. The MCWCRC at Boston Medical Center will seek to overcome this trend by showing that a preventative, embedded psychiatric approach will lead to cost savings in the long term, not in the least part by improving the current 50% no-show rate, capturing patients with nascent symptoms early on, and increasing utilization of clinic services by providing screening and education on the spot. We anticipate that the use of our mobile platform, which includes the iPad behavioral health screen and the mobile phone app aimed at education and care coordination, will prevent patients from seeking help in the emergency

department for conditions that can be successfully addressed within the embedded psychiatric clinic; in the long run, this could lead to cost savings for the medical center, the taxpayer, and the patient.

Figure 3. Visual representation of the patient flow during the pilot phase of the mHealth prospective cohort study in the Mother-Child Wellness Clinical and Research Center at Boston Medical Center. AC-OK: Andrew Cherry-Oklahoma Co-Occurring Disorders Screen; EPDS: Edinburgh Postnatal Depression Scale; LEC-5: Life Events Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; Ob: obstetric; Psych: psychiatric; Ob/Gyn: obstetrician-gynecologist; PPD: postpartum depression.



Conclusions

Prior studies have shown that mHealth apps have the potential to improve outcomes among patients in OB-GYN settings; further research is needed to evaluate whether these benefits extend to obstetric clinics with embedded psychiatric care. The wealth of information available from global health research indicates a high potential for improved mental and reproductive health outcomes. However, limited research has been conducted in the United States to determine the health outcome benefits associated with mHealth, particularly in a low socioeconomic status, inner city population. Furthermore, there is no comprehensive app on the US market that incorporates patient education, behavior change, mood tracking, appointment reminders, and community engagement. The MCWCRC at Boston Medical Center will endeavor to fill this gap with the upcoming mHealth app trial.

Acknowledgments

The authors thank the staff of the Boston Medical Center OB-GYN Department as well as the Boston University School of Medicine staff and faculty for their support and assistance.

CPCB received a Mentored Career Development Award (K01MH100428) from the National Institutes of Health. SM received support from the Department of Psychiatry, Boston Medical Center, and Boston University School of Medicine.

Conflicts of Interest

None declared.

References

1. Kessler RC, Birnbaum H, Bromet E, Hwang I, Sampson N, Shahly V. Age differences in major depression: Results from the National Comorbidity Survey Replication (NCS-R). Psychol Med 2010 Feb;40(2):225-237 [FREE Full text] [doi: 10.1017/S0033291709990213] [Medline: 19531277]
2. Dennis C, Chung-Lee L. Postpartum depression help-seeking barriers and maternal treatment preferences: A qualitative systematic review. Birth 2006 Dec;33(4):323-331. [doi: 10.1111/j.1523-536X.2006.00130.x] [Medline: 17150072]
3. Leddy MA, Lawrence H, Schulkin J. Obstetrician-gynecologists and women's mental health: Findings of the Collaborative Ambulatory Research Network 2005-2009. Obstet Gynecol Surv 2011 May;66(5):316-323. [doi: 10.1097/OGX.0b013e31822785ee] [Medline: 21794195]
4. Scholle SH, Chang JC, Harman J, McNeil M. Trends in women's health services by type of physician seen: Data from the 1985 and 1997-98 NAMCS. Womens Health Issues 2002;12(4):165-177. [Medline: 12093581]
5. Miranda J, Azocar F, Komaromy M, Golding JM. Unmet mental health needs of women in public-sector gynecologic clinics. Am J Obstet Gynecol 1998 Feb;178(2):212-217. [Medline: 9500476]
6. Melville JL, Reed SD, Russo J, Croicu CA, Ludman E, LaRocco-Cockburn A, et al. Improving care for depression in obstetrics and gynecology: A randomized controlled trial. Obstet Gynecol 2014 Jun;123(6):1237-1246 [FREE Full text] [doi: 10.1097/AOG.0000000000000231] [Medline: 24807320]
7. Poleshuck EL, Woods J. Psychologists partnering with obstetricians and gynecologists: Meeting the need for patient-centered models of women's health care delivery. Am Psychol 2014;69(4):344-354. [doi: 10.1037/a0036044] [Medline: 24820684]
8. Katon W, Russo J, Reed SD, Croicu CA, Ludman E, LaRocco A, et al. A randomized trial of collaborative depression care in obstetrics and gynecology clinics: Socioeconomic disadvantage and treatment response. Am J Psychiatry 2015 Jan;172(1):32-40 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14020258] [Medline: 25157500]
9. Melek SP, Norris DT, Paulus J. Economic Impact of Integrated Medical-Behavioral Healthcare: Implications for Psychiatry. Denver, CO: Milliman, Inc; 2014 Apr. URL: https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/Professional-Topics/Integrated-Care/Milliman-Report-Economic-Impact-Integrated-Implications-Psychiatry.pdf [accessed 2017-09-21] [WebCite Cache ID 6te9s3CUJ]
10. World Health Organization. World Health Organization. Geneva, Switzerland: World Health Organization; 2011. mHealth: New horizons for health through mobile technologies URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2017-03-13] [WebCite Cache ID 6owQBAltY]
11. Adibi S, editor. Mobile Health: A Technology Road Map. Cham, Switzerland: Springer International Publishing; 2015.
12. Ericsson. 2013 Nov 13. Ericsson Mobility Report: Global smartphone subscriptions to reach 5.6 billion by 2019 URL: https://www.ericsson.com/news/1741771 [accessed 2017-03-13] [WebCite Cache ID 6owRhPjko]
13. Perera C. The evolution of eHealth: Mobile technology and mHealth. J Mob Technol Med 2012 Mar;1(1):1-2. [doi: 10.7309/jmtm11]
14. Technopedia. Mobile application (mobile app) URL: https://www.techopedia.com/definition/2953/mobile-application-mobile-app [accessed 2017-03-13] [WebCite Cache ID 6owQjEIoO]
15. Maternal mHealth: Solution/Product and Technology Framework. London, UK: GSMA URL: http://www.gsma.com/mobilefordevelopment/wp-content/uploads/2014/12/MOBILE-MATERNAL-HEALTH-PRODUCT-AND-TECHNOLOGY-FRAMEWORK-STYLED-v16_tiffs.pdf [accessed 2017-03-13] [WebCite Cache ID 6owRDV8oT]
16. Pennic F. HIT Consultant. 2012 Nov 28. State of mHealth: Present, adoption, economics, barriers, & future infographic URL: http://hitconsultant.net/2012/11/28/state-of-mhealth-present-adoption-economics-barriers-future/ [accessed 2017-03-13] [WebCite Cache ID 6owRPAFaO]
17. 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]
18. 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]
19. United Nations Foundation. Mobile Alliance for Maternal Action (MAMA) URL: http://www.mobilemamaalliance.org/ [accessed 2017-03-13] [WebCite Cache ID 6owRVYFPO]
20. Odetola TD. Conceptual model for mHealth utilization: A Nigerian adaptation. iProc 2015 Oct 27;1(1):e3. [doi: 10.2196/iproc.4645]
21. 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]
22. Ngabo F, Nguimfack J, Nwaigwe F, Mugeni C, Muhoza D, Wilson DR, et al. Designing and implementing an innovative SMS-based alert system (RapidSMS-MCH) to monitor pregnancy and reduce maternal and child deaths in Rwanda. Pan Afr Med J 2012;13:31 [FREE Full text] [Medline: 23330022]

23. Luxton D, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: Integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011;42(6):505-512. [doi: 10.1037/a0024485]
24. Alvidrez J, Azocar F. Distressed women's clinic patients: Preferences for mental health treatments and perceived obstacles. Gen Hosp Psychiatry 1999;21(5):340-347. [Medline: 10572775]
25. CliniOps. URL: http://cliniops.com [accessed 2017-03-13] [WebCite Cache ID 6owSvwIYl]
26. Cox J, Holden J. Perinatal Mental Health: A Guide to the Edinburgh Postnatal Depression Scale. London, UK: Gaskell; 2003.
27. Weathers F, Blake D, Schnurr P, Kaloupek D, Marx B, Keane T. National Center for PTSD. 2013. Life Events Checklist for DSM-5 (LEC-5) URL: https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp [accessed 2017-09-19] [WebCite Cache ID 6ta9s9cAD]
28. Cherry AL, Dillon ME. The AC-OK Cooccurring Screen: Reliability, convergent validity, sensitivity, and specificity. J Addict 2013;2013:573906 [FREE Full text] [doi: 10.1155/2013/573906] [Medline: 24826362]
29. Anderson M. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2015 Oct 29. Technology device ownership: 2015 URL: http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/ [accessed 2017-03-13] [WebCite Cache ID 6owTG8sJM]
30. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: Implications for public health practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
31. 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]

Abbreviations

AC-OK COD: Andrew Cherry-Oklahoma Co-Occurring Disorders
EPDS: Edinburgh Postnatal Depression Scale
LEC-5: Life Events Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
MAMA: Mobile Alliance for Maternal Action
MCWCRC: Mother-Child Wellness Clinical and Research Center
mHealth: mobile health
OB-GYN: obstetrician-gynecologist
PPD ACT: Postpartum Depression: Action Towards Causes and Treatment
SMS: short message service

Edited by G Eysenbach; submitted 26.05.17; peer-reviewed by J Payne, D Linares, HN Le, S Shorey; comments to author 15.06.17; revised version received 10.08.17; accepted 28.08.17; published 06.10.17.

Please cite as:
Mehralizade A, Schor S, Coleman CM, Oppenheim CE, Denckla CA, Borba CPC, Henderson DC, Wolff J, Crane S, Nettles-Gomez P, Pal A, Milanovic S
Mobile Health Apps in OB-GYN-Embedded Psychiatric Care: Commentary
JMIR Mhealth Uhealth 2017;5(10):e152
URL: http://mhealth.jmir.org/2017/10/e152/
doi:10.2196/mhealth.7988
PMID:28986335

Original Paper

Using Google Glass in Nonsurgical Medical Settings: Systematic Review

Bryn Dougherty[1], BS; Sherif M Badawy[2,3,4], MD, MS, MBBCh

[1]Northwestern University Weinberg College of Arts and Sciences, Evanston, IL, United States

[2]Division of Hematology, Oncology and Stem Cell Transplantation, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[3]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[4]Department of Pediatrics, Division of Hematology and Oncology, Zagazig University Faculty of Medicine, Zagazig, Egypt

Corresponding Author:
Sherif M Badawy, MD, MS, MBBCh
Division of Hematology, Oncology and Stem Cell Transplantation
Ann & Robert H. Lurie Children's Hospital of Chicago
225 E. Chicago Ave, Box #30
Chicago, IL, 60611
United States
Phone: 1 312 227 4836
Fax: 1 312 227 9376
Email: sbadawy@luriechildrens.org

Abstract

Background: Wearable technologies provide users hands-free access to computer functions and are becoming increasingly popular on both the consumer market and in various industries. The medical industry has pioneered research and implementation of head-mounted wearable devices, such as Google Glass. Most of this research has focused on surgical interventions; however, other medical fields have begun to explore the potential of this technology to support both patients and clinicians.

Objective: Our aim was to systematically evaluate the feasibility, usability, and acceptability of using Google Glass in nonsurgical medical settings and to determine the benefits, limitations, and future directions of its application.

Methods: This review covers literature published between January 2013 and May 2017. Searches included PubMed MEDLINE, Embase, INSPEC (Ebsco), Cochrane Central Register of Controlled Trials (CENTRAL), IEEE Explore, Web of Science, Scopus, and Compendex. The search strategy sought all articles on Google Glass. Two reviewers independently screened titles and abstracts, assessed full-text articles, and extracted data from articles that met all predefined criteria. Any disagreements were resolved by discussion or consultation by the senior author. Included studies were original research articles that evaluated the feasibility, usability, or acceptability of Google Glass in nonsurgical medical settings. The preferred reporting results of systematic reviews and meta-analyses (PRISMA) guidelines were followed for reporting of results.

Results: Of the 852 records examined, 51 met all predefined criteria, including patient-centered (n=21) and clinician-centered studies (n=30). Patient-centered studies explored the utility of Google Glass in supporting patients with motor impairments (n=8), visual impairments (n=5), developmental and psychiatric disorders (n=2), weight management concerns (n=3), allergies (n=1), or other health concerns (n=2). Clinician-centered studies explored the utility of Google Glass in student training (n=9), disaster relief (n=4), diagnostics (n=2), nursing (n=1), autopsy and postmortem examination (n=1), wound care (n=1), behavioral sciences (n=1), and various medical subspecialties, including, cardiology (n=3), radiology (n=3), neurology (n=1), anesthesiology (n=1), pulmonology (n=1), toxicology (n=1), and dermatology (n=1). Most of the studies were conducted in the United States (40/51, 78%), did not report specific age information for participants (38/51, 75%), had sample size <30 participants (29/51, 57%), and were pilot or feasibility studies (31/51, 61%). Most patient-centered studies (19/21, 90%) demonstrated feasibility with high satisfaction and acceptability among participants, despite a few technical challenges with the device. A number of clinician-centered studies (11/30, 37%) reported low to moderate satisfaction among participants, with the most promising results being in the area of student training. Studies varied in sample size, approach for implementation of Google Glass, and outcomes assessment.

Conclusions: The use of Google Glass in nonsurgical medical settings varied. More promising results regarding the feasibility, usability, and acceptability of using Google Glass were seen in patient-centered studies and student training settings. Further research evaluating the efficacy and cost-effectiveness of Google Glass as an intervention to improve important clinical outcomes is warranted.

(JMIR Mhealth Uhealth 2017;5(10):e159) doi:10.2196/mhealth.8671

KEYWORDS

Google Glass; wearable; wearable device; head-mounted wearable device; non-surgical setting; non-surgical condition; medical setting; medical condition

Introduction

Wearable technology is defined as any compact device, either in the form of a body sensor or head-mounted display, which provides a user information and allows user interaction via voice command or physical input [1]. The purpose of these devices is to create convenient, portable, and hands-free access to computers, thus facilitating or enhancing everyday tasks. Many of these devices can perform the same functions as mobile phones and laptop computers, while also outperforming them with their sensory and scanning abilities [2]. Google Glass (Google, Inc.), often referred to as "Glass," which resembles standard eyeglasses, is one of the more well-known devices in this emerging field since its release in 2013 [3].

Google Glass has distinguished itself from other head-mounted or heads-up wearable devices by providing users with a comfortable, unobtrusive, wireless platform that runs the Android operating system and displays virtual or augmented reality with little obstruction to normal vision [3]. While it has not yet seen much success in the consumer market, various industries have taken an interest in the potential applications of a head-mounted, ubiquitous computer that could be used for a range of tasks, including recording and streaming videos, data transmission, telementoring in education, and teleconferences for professional collaboration [3]. Health care is one such industry that has pioneered research investigating how Google Glass could be leveraged to support both clinicians and patients.

Surgeons were among the first in the medical industry to incorporate Google Glass into their work. As a hands-free device that can react to voice commands, eye movements, and simple gestures, it is particularly attractive in environments where both hands are generally occupied with surgical tasks and maintaining sterility is of upmost importance [4]. In a recent systematic review, Davis and Rosenfield reported an overall positive impact of using Google Glass in surgical settings with data to support the feasibility and acceptability of its use for medical care, surgical skills training, medical documentation, and patient safety [4]. Many other specialties in medicine have followed the lead of the surgical field and conducted their own studies to assess the feasibility of using Google Glass in nonsurgical medical settings.

While Google Glass is an exciting technology with a number of promising applications in medicine, it remains unclear which applications are most worth pursuing, what potential limitations are associated with its use, and the extent to which patients and clinicians might benefit from its use. The objectives of this review are to systematically evaluate the most recent evidence for the feasibility, usability, and acceptability of using Google Glass in nonsurgical settings, and determine its potential benefits, limitations, and future directions in these settings.

Methods

We followed the guidelines for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) in the reporting of evidence across the studies we reviewed (Multimedia Appendix 1) [5].

Article Retrieval

A librarian collaboratively developed the search strategies with the senior author (SB) and ran searches in the following databases in November 2015: PubMed MEDLINE, Embase, INSPEC (Ebsco), Cochrane Central Register of Controlled Trials (CENTRAL) on the Wiley platform, IEEE Explore, Web of Science, Scopus, and Compendex. An updated search of all databases was run in January 2017 to look for additional articles. Search strategies for all databases except MEDLINE were adapted from the PubMed MEDLINE strategy. All databases were searched back to 2013, when Google Glass was first released. No language limits were applied. The search strategy specified keywords related to Google Glass. We also reviewed the search strategies of previous studies to include additional terms. See Multimedia Appendix 2 for complete search strategies in each database. An additional hand-search of related themes in the *Journal of Medical Internet Research* was also conducted. We also attempted to identify additional studies by searching the reference lists of key studies and relevant systematic reviews.

Study Selection

The inclusion criteria were as follows: (1) original research articles, (2) studies that were either randomized controlled trials, quasi-experimental studies, or pilot/feasibility studies (including single arm, pre-posttest), (3) Google Glass interventions, (4) nonsurgical study settings, and (5) clinical, usability, feasibility, and/or acceptability as primary or secondary outcome. The exclusion criteria included (1) technology-based interventions other than Google Glass, (2) surgical study settings, and (3) articles with more technical description of Google Glass but no clinical, usability, feasibility, and/or acceptability outcomes.

Data Extraction and Analysis

We used a standardized form for data extraction. Data items in the extraction form included the following: first author's name, publication year, country, condition or disease focus of the study, purpose of the study, description of how Google Glass was used in the study as an intervention, participants' age (when available), study design, study setting, duration of the study, and other study considerations. Two authors coded all included articles individually. Disagreements were resolved by discussion or by consultation with the senior author (SB), if needed. Quantitative and qualitative data analyses were conducted.

Results

Literature Search

The literature search identified 852 references (Figure 1), and 498 individual full articles were retrieved. A total of 51 articles met all inclusion criteria. Some of the interventions (21/51, 41%) studied the potential of Google Glass in aiding patients with a variety of conditions [6-26], while the majority (30/51, 59%) studied its potential uses in assisting health care professionals in their work [27-56]. The patient-focused studies aimed to help individuals with motor impairments (8/21, 38%) [6-13], visual impairments (5/21, 24%) [14-18], developmental and psychiatric disorders (2/21, 9%) [19,20], weight management concerns (3/21, 14%) [21-23], allergies (1/21, 5%) [26], or other health concerns leading them to track specific physiological metrics (2/21, 10%) [24,25]. The clinician-focused studies analyzed Google Glass use in student training (9/30, 30%) [35,37,38,40,41,46,48,49,52], disaster relief (4/30, 13%) [27-30], diagnostics (2/30, 7%) [32,50], nursing (1/30, 3%) [33], autopsy and postmortem examination (1/30, 3%) [53], wound care (1/30, 3%) [54], behavioral sciences (1/30, 3%) [31], and various medical specialties, including cardiology (3/30, 10%) [43-45], radiology (3/30, 10%) [39,42,47], neurology (1/30, 3%) [34], anesthesiology (1/30, 3%) [36], pulmonology (1/30, 3%) [51], toxicology (1/30, 3%) [55], and dermatology (1/30, 3%) [56].

Description of Included Studies

Tables 1 and 2 summarize the characteristics of patient- and clinician-centered studies, respectively. In total, 40 studies were conducted in the United States [6-9,13-18,20-24,27,29,31,32,35,37-41,43-52,54-57], three in Germany [25,26,53], two in United Kingdom [10,11], China [34,36], and one each in Australia [33], Switzerland [42], Mexico [19], Netherlands [12], Norway [30], and Italy [28]. Less than half of the included studies (19/51, 37%) were conducted in a laboratory setting [6-9,12,14,16-19,21-24,27,31,33,42,50], 13 (25%) in a hospital setting [32,34,36,38,39,43,44,46,51,53-56], seven (14%) in a classroom or clinical student training setting [35,37,40,41,47,48,52], three in patient residences (6%) [13,15,20], three in local settings (6%) [28-30], and one in a dental office (2%) [49]. The remaining five studies were conducted in varying locations (10%) [10,11,25,26,45]. There was significant variability in information reported about participant demographics. Most (n=38) did not report any specific age information for participants [7,8,13-15,19-21,26-52,54-56], but none of these were conducted in pediatric settings. Of the 13 studies that did report participant age information, seven enrolled young adults (average age or age range ≤35 years) [6,18,22-25,53], three enrolled adults (average age or age range >36 and <60 years) [9,16,17], two enrolled older adults (average age or age range ≥60 years) [10,12] and one study reported an age range of 46-70 years [11]. Sample size ranged from 1-106 participants, with a median of 12 and a mean of 22 participants per study; 29 enrolled <30 [6,7,9-14,16-21,23-27,30,38-40,43-45,48,53,54] and 10 had ≥30 participants [22,31,33,36,37,41,47,49,52,56]. Some of the studies (12/51, 23.5%) did not report the number of participants [8,15,28,29,32,34,35,42,46,50,51,55]. None of the studies reported information about participants' race and ethnicity. Most (31/51, 61%) were pilot or feasibility studies [6-16,18,20-22, 24,26-29,32,33,35,36,38,47,48,53-56], six were randomized controlled trials (6/51, 12%) [30,31,37,40,41,52], five were exploratory studies (5/51, 10%) [42,44,45,49,50], five were case studies (5/51, 10%) [17,23,34,46,51], and four were quasi-experimental (4/51, 8%) [19,25,39,43]. None of the studies included any follow-up with participants after completion of the intervention.

Table 1. Summary of studies using Google Glass as patient-centered interventions.

Source (country)	Health condition	Study design	Study setting	Google Glass (GG) use
Anam et al, 2014 (United States)	Ophthalmology – visual impairment	Pilot/feasibility study	Laboratory	Monitors and reports nonverbal social cues to user
Garcia and Nahapetian, 2015 (United States)	Ophthalmology – visual impairment	Pilot/feasibility study	Patient home	Analyzes environment and reports the information to user to help them navigate a room
Pundlik et al, 2016 (United States)	Ophthalmology – visual impairment	Pilot/feasibility study	Laboratory	Magnifies user's vision while completing a series of tasks
Hwang and Peli, 2016 (United States)	Ophthalmology – advanced age-related macular degeneration	Case study	Laboratory	Warps the vision of participants in efforts to improve vision
Tanuwidjaja et al, 2014 (United States)	Ophthalmology – color-blindness	Pilot/feasibility study	Laboratory	Helps participants identify colors
Lazewatsky et al, 2014 (United States)	Motor impairment	Pilot/feasibility study	Laboratory	Helps participants guide the robot personal assistant
Gips et al, 2015 (United States)	Motor impairment	Pilot/feasibility study	Laboratory	Allows people to operate a computer with only eye or head movements
Sinyukov et al, 2016 (United States)	Motor impairment – Locked-In Syndrome	Pilot/feasibility study	Laboratory	Uses voice control function of GG to allow people to navigate an electric wheelchair in indoor environments
Malu and Findlater, 2015 (United States)	Motor impairment – upper body impairment	Pilot/feasibility study	Laboratory	Uses touchpad and visual display to perform tasks on a computer/ mobile phone
McNaney et al, 2014 (United Kingdom)	Motor impairment – Parkinson's Disease	Pilot/feasibility study	Varying locations (patient home, in public)	Helps in daily interactions and common activities
McNaney et al, 2015 (United Kingdom)	Motor impairment – Parkinson's Disease	Pilot/feasibility study	Varying locations (patient home, in public)	Monitors user's speech volume and provides feedback
Zhao et al, 2016 (Netherlands)	Motor impairment – Parkinson's Disease	Pilot/feasibility study	Laboratory	Provides visual and auditory cues to modulate gait
Pervaiz and Patel, 2014 (United States)	Motor impairment – Dysarthria	Pilot/feasibility study	Assisted living facility	Helps people be aware of their volume, notifies them when to raise it, and provides feedback to clinicians so they can adjust therapy
Miranda et al, 2014 (Mexico)	Psychiatric/Developmental – Social Anxiety Disorder (SAD)	Quasi-experimental	Laboratory	Monitors symptoms of SAD through blinking habits
Voss et al, 2016 (United States)	Psychiatric/Developmental – Children with Autism Spectrum Disorder (ASD)	Pilot/feasibility study	Patient home	Uses the video feature to monitor everyday life
Mirtchouk et al, 2016 (United States)	Eating monitoring	Pilot/feasibility study	Laboratory	Records head motion while participants eat
Rahman et al, 2015 (United States)	Eating monitoring	Pilot/feasibility study	Laboratory	Records user's eating and drinking habits through head movements
Ye et al, 2015 (United States)	Eating monitoring	Case study	Laboratory	Records head motion while participants eat
Hernandez et al, 2014 (United States)	Physiological measurements	Pilot/feasibility study	Laboratory	The accelerometer, gyroscope, and camera on GG are used to analyze the heart and respiration rate of user wearing the device
Richer et al, 2015 (Germany)	Physiological measurements	Quasi-experimental study	Varying locations (patients' everyday lives)	Serves as the "wearable extension" portion of the DailyHeart app

Source (country)	Health condition	Study design	Study setting	Google Glass (GG) use
Wiesner et al, 2015 (Germany)	Allergies	Pilot/feasibility study	Varying locations (drugstores selling cosmetic products)	Cross checks ingredients on cosmetic product package with a list of allergens created by the user in their online profile

Description of Google Glass Use as Patient-Centered Interventions

Table 3 summarizes the Google Glass approach as patient-centered interventions. Five of the studies (5/21, 24%) used Google Glass to assist individuals with visual impairments or in low vision environments by providing them information about nonverbal social cues [14], allowing them to better navigate environments with the use of floor plans [15], improving vision magnification using mobile phone zoom capabilities [16], compensating for age-related vision impairments [17], and augmenting color perception [18]. Eight studies (8/21, 38%) using Google Glass to help individuals with various motor impairments provided them with an accessible interface to control an assistive robot [6] or an electric wheelchair [8]. This allowed them to operate a computer using only head or eye movements [7], facilitating everyday tasks with the use of voice commands and the touchpad [9], managing symptoms of Parkinson's Disease (PD) [10-12], and providing speech feedback to patients with dysarthria to allow them to better adjust their volume [13]. Two studies (2/21, 10%) used Google Glass to help individuals with psychiatric or developmental disorders by recording blinking information as an indication of anxiety experienced by those with social anxiety disorder (SAD) [19] and by recording behaviors of individuals with autistic spectrum disorder (ASD) to provide better information to caregivers and clinicians [20]. Three studies (3/21, 14%) used Google Glass to assist in weight management by detecting and recording a person's eating and drinking habits [21-23]. Two studies (2/21, 10%) provided individuals with real-time electrocardiograms (ECG) [25] or other physiological measurement feedback [24]. Finally, one study (1/21, 5%) allowed users to scan the ingredients of cosmetic products in drug stores to filter for common allergens [26].

Figure 1. Flow of studies according to PRISMA guidelines.

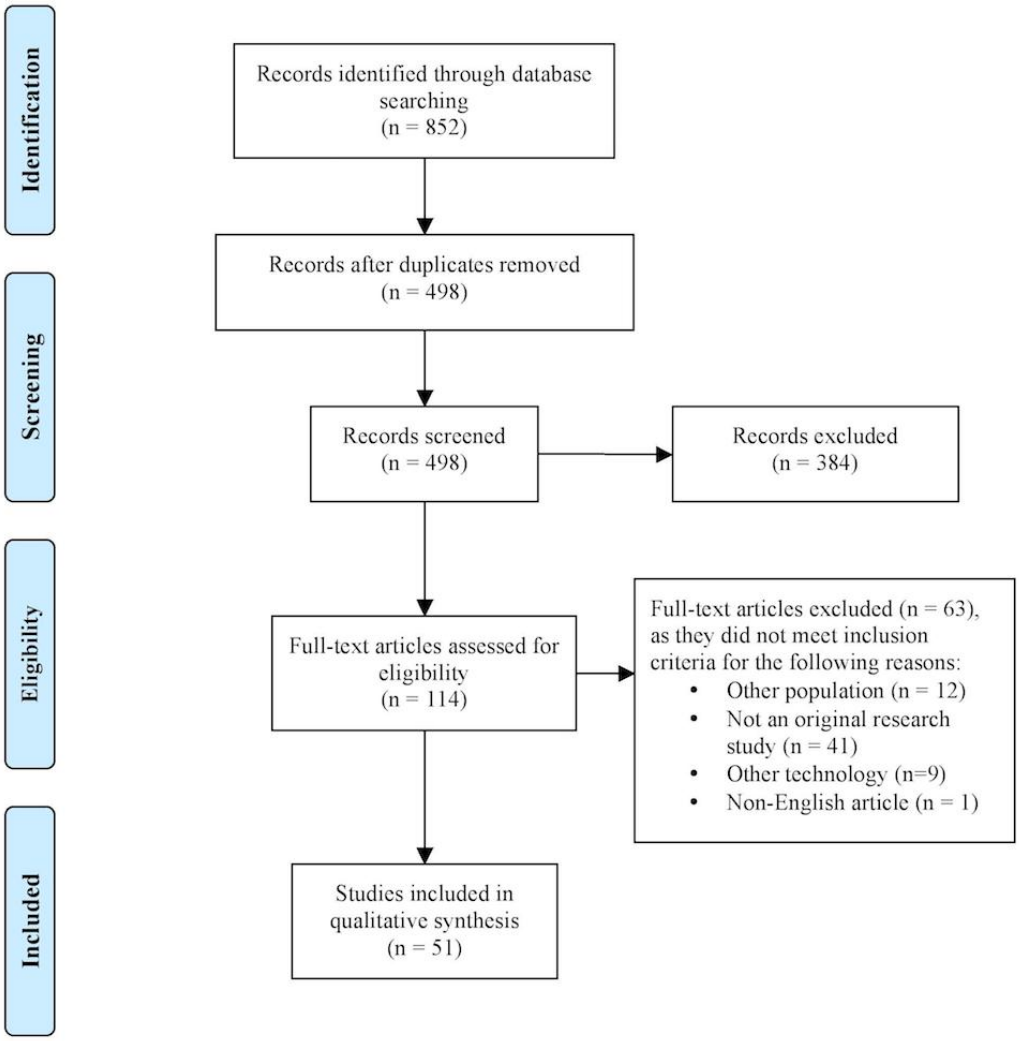


Table 2. Summary of studies using Google Glass as clinician-focused interventions.

Source (country)	Health condition	Study design	Study setting	Google Glass (GG) use
Gillis et al, 2015 (United States)	Disaster relief	Pilot/feasibility study	Laboratory	Allows for audiovisual communication with each group of paramedics and the administrator; the virtual beacon component is used to eliminate the use for paper triage tags
Carenzo et al, 2014 (Italy)	Disaster relief	Pilot/feasibility study	Local – Field hospital	Scans triage tags to provide their information, timestamp, and Global Positioning System (GPS) coordinates and relay the information back to the hospital
Cicero et al, 2014 (United States)	Disaster relief	Pilot/feasibility study	Local – Airport	Facilitates communication with telemedicine physician disaster expert who can confirm the triage decision of the intervention team, and determines time of triage for each patient
Newaz and Eide, 2015 (Norway)	Disaster relief	Randomized control trial	Local – Neighborhood	Provides navigation and maps to first responders
Paxton et al, 2015 (United States)	Behavioral sciences	Randomized control trial	Laboratory	Uses PsyGlass app to facilitate the use of GG in behavioral, cognitive, and social research
Pappachan et al, 2014 (United States)	Diagnostics	Pilot/feasibility study	Hospital – Emergency department	Helps community health workers to identify certain disorders based on the patient demographics
Pascale et al, 2015 (Australia)	Nursing – Peripheral detection	Pilot/feasibility study	Laboratory	Improves detection in the periphery
Yuan et al, 2015 (China)	Neurology	Case study	Hospital – Neurology	Facilitates communication between physicians during a neurological examination
Chaballout et al, 2016 (United States)	Student training – health science students	Pilot/feasibility study	Classroom (university)	Presents a simulation in conjunction with real time performance of treatment on a manikin
Drake-Brockman et al, 2016 (China)	Anesthesiology	Pilot/feasibility study	Hospital – Anesthesiology	Uses heads-up display to facilitate monitoring patient vitals while performing procedures
Iversen et al, 2016 (United States)	Student training – physiotherapy students	Randomized control trial	Classroom (large, private, non-profit research university)	Captures 1st-person view of a procedure and displays it for learning purposes
Son et al, 2015 (United States)	Student training – otolaryngology residents	Pilot/feasibility study	Hospital - Otolaryngology	Uses the video capabilities to record resident encounters with patients
Spaedy et al, 2016 (United States)	Radiology	Quasi-experimental	Hospital – Radiology	Takes images of and displays X-rays for physician interpretation
Russel et al, 2014 (United States)	Student training – medical students (radiology)	Randomized controlled trial	Instructional testing room (University of Kentucky, School of Medicine)	Provides live instruction from an expert via Google Hangout
Wu et al, 2014 (United States)	Student training – medical students and radiology residents	Randomized controlled trial	Classroom (University of Arizona College of Medicine – Phoenix)	Facilitates procedures by showing real-time ultrasound images on the heads-up display
Widmer et al, 2014 (Switzerland)	Dermatology and Radiology	Exploratory study	Laboratory	Takes and analyzes images to facilitate interpretation and diagnostic decisions by presenting similar images to user
Stetler et al, 2015 (United States)	Cardiology	Quasi-experimental	Hospital – Cardiology	Captures images of the electrocardiogram (ECGs) and presents them on heads-up display to facilitate interpretation
Duong et al, 2015 (United States)	Cardiology	Exploratory study	Hospital – Cardiology	Records coronary angiograms were recorded to be reviewed on the heads-up display or transferred to a mobile phone

Source (country)	Health condition	Study design	Study setting	Google Glass (GG) use
Jeroudi et al, 2014 (United States)	Cardiology	Exploratory study	Remote (location varied by physician reviewer)	Displays ECG images for interpretation
Vallurupalli et al, 2013 (United States)	Student training – medical students (cardiology)	Case study	Hospital – Cardiology (University of Arkansas for Medical Sciences, Division of Cardiology)	Facilitates collaboration between cardiology attending and resident in clinical training settings
Benninger, 2015 (United States)	Radiology	Pilot/feasibility study	Anatomy laboratory (Medical School)	Displays images captured by an ultrasound finger probe to teach medical students anatomy and simple interventions
Vaughn et al, 2016 (United States)	Student training – nursing students	Pilot/feasibility study	Classroom (Nursing School)	Presents a simulation in conjunction with real-time performance of treatment on a manikin
Zahl et al, 2016 (United States)	Student training – dental students	Exploratory study	Dental office	Records student SP station for later analysis
Feng et al, 2015 (United States)	Diagnostics – Human immunodeficiency virus (HIV) and cancer	Exploratory study	Laboratory	Takes rapid diagnostic tests (RDTs), images prostate specific antigen (PSA) tests, and images previously activated free PSA and total PSA RDTs
Spencer et al, 2014 (United States)	Pulmonology – airway assessment for burn victims	Case study	Hospital – Burn unit (Massachusetts General Hospital)	Facilitates assessment and management of the airway
Tully et al, 2015 (United States)	Student training – medical students (hospice)	Randomized controlled trial	University of Arizona, College of Medicine: Phoenix and local hospice organization	Records student standardized patient encounters for later analysis
Albrecht et al, 2014 (Germany)	Pathology – autopsy and postmortem examinations	Pilot/feasibility study	Hospital – Autopsy laboratory	Takes pictures of body for documentation during examination
Aldaz et al, 2015 (United States)	Chronic wounds	Pilot/feasibility study	Hospital – Wound care (Stanford Hospital and Clinics)	Uses SnapCap software to facilitate hands-free digital imaging and the tagging and transfer of images to patient's electronic medical record in chronic wound care assessments
Chai et al, 2015 (United States)	Toxicology	Pilot/feasibility study	Hospital – Emergency department (urban academic hospital)	Sends photographs and videos to the toxicology supervisors; acts as a platform for instruction of 2nd-year medical staff
Chai et al, 2015 (United States)	Dermatology	Pilot/feasibility study	Hospital – Emergency department (urban academic hospital)	Allows teledermatolgosists to complete a dermatology assessment via live video feed after in-person consultation by a resident

Table 3. Summary of Google Glass approach as patient-centered interventions.

Source (health condition)	Purpose	Intervention description
Anam et al, 2014 (Ophthalmology – visual impairment)	To allow people with vision impairments gain the ability to determine non-verbal expressions	Expression is the type of feature addition that is being used
		It analyzes changes in facial expression and relays that information in the form of captured frames to user
		Helps user change their posture to better capture the facial expression
Garcia and Nahapetian, 2015 (Ophthalmology – visual impairment)	To help guide people with visual impairments navigate indoor environments	Extract floor regions from images captured from GG to help guide the individual
		An app is installed in GG that starts the camera and sends image frames to the mobile phone
		An app is also installed that analyzes the floor plans and then sends it to the mobile phone through Bluetooth
		Images that are captured contain the walls, floor, and ceiling
Pundlik et al, 2016 (Ophthalmology – visual impairment)	To use vision magnification to aid in the completion of tasks	Leverages zoom capabilities of GG
		Students are assigned tasks that involve the calculator and music player apps
		Performance on these tasks is measured
Hwang and Peli, 2016 (Ophthalmology – advanced age-related macular degeneration)	To augment the vision of the wearer so that they have improved vision	Vision enhancement tool is added to GG
		Participant wears GG which now warps the camera image to improve vision
		Images that the vision enhancement tool sees are then relayed to user in real-time
Tanuwidjaja et al, 2014 (Ophthalmology – colorblindness)	To help people with colorblindness see color	Alters the way people perceive color
		Applied Chroma, which is an app that detects color and relays that information to the participant
		Implemented the Ishihara test, which tests for color vision deficiency
		Implemented the Blackboard test that determines if a person can distinguish between green and orange
Lazewatsky et al, 2014 (Motor impairment)	To show that GG can be used in conjunction with the PR2 robot to recognize people and objects and then manipulate the space around it	GG Bridge Node receives sensor data from GG and transmits it to Robots and Systems software (ROS) messages and publishes a coordinate frame for GG
		ROS works with face detection; GG software also uses face detection and person recognition
Gips et al, 2015 (Motor impairment)	To help people operate a computer with only eye or head movements	Noggin software was developed to allow user to move a cursor across the screen through head movements
		Noggin displays yes, no, and enter on the screen
		Noggin uses the gyroscope to monitor head movements
		GG Gab, another software, allows user to spell out a message
Sinyukov et al, 2016 (Motor impairment – Locked-In Syndrome)	To help patients have better control over their wheelchairs	Patient uses the software installed on GG in conjunction with the motorized wheelchair
		GG monitors facial expressions of the patient
		GG's audio monitoring is used to understand voice commands and then relay the instructions to the motorized wheelchair

Source (health condition)	Purpose	Intervention description
Malu and Findlater, 2015 (Motor impairment – upper body)	To assess the accessibility of GG for individuals with upper body motor impairments	Using voice commands and the touchpad to go through day-to-day activities
		Touchpad on GG was on the right arm of the device and senses taps and swipes through voice commands
		Output is projected on the heads-up display
		Participants completed tasks using swipes and tasks function
		Participants then used a scale to rate the comport and ease of the touchpad and visual display
McNaney et al, 2014 (Motor impairment – Parkinson's Disease [PD])	To help people with PD counteract their symptoms by allowing them to carry out the normal functions of a mobile phone using voice commands, cueing for freezing gait	GG was used to manage social cues and alert the user
		GG monitored movement and told the participant when they were freezing so that they could actively try to stop the behavior
McNaney et al, 2015 (Motor impairment – PD)	To help monitor speech loudness issues and provide feedback to help with self-management	Developed the LApp app that monitors loudness
		Participants used the app for a set amount of time while carrying out a series of social interactions
		Indicating when the volume was inappropriate so the user could adjust to hit the target loudness
Zhao et al, 2016 (Motor impairment – PD)	To provide visual and auditory cues to aid in the modulation of gait	GG was used to detect gait issues and improve them through cueing
		Audiovisual cues were used, including a metronome, flashing light, optic flow, and a control (no cue)
		Participants underwent a series of walking tasks and their gait was then analyzed for stability and freezing
Pervaiz and Patel, 2014 (Motor impairment – Dysarthria)	To help patients monitor their low volume in order to self-regulate and to provide clinicians with feedback to adjust therapy	Developed the SpeedOmeter software that compares vocal loudness to ambient noise
		Provides feedback to user on their volume
		System provides usage and performance history for user
		Notifies patient of their volume so they can adjust
Miranda et al, 2014 (Psychological/Developmental – SAD)	To assess the feasibility of using GG to monitor blinking rates in individuals with social anxiety disorder	Monitor blinking behaviors
		Used to gather data from the infrared (IR) sensor
		The app dealt with IR data gathering, data processing, and HTTP communication
		App processes the data and calculates when the user blinked
Voss et al, 2016 (Psychological/Developmental – ASD)	To monitor life activities and allow for analysis of autism behaviors	Participant uses GG to record everyday behaviors
		Caregiver reviews system highlights and emotional moments so they are easily accessible for the reviewer
		Caregivers can tag parts of the video that are especially important and add comments to the video
Mirtchouk et al, 2016 (Eating monitoring)	To accurately track an individual's eating habits and provide feedback to help with self-regulation	GG sensor was used to detect head movement that was specific to eating
		Participants ate what they wanted and when they wanted and GG was supposed to detect when they were eating and for how long
		Participants were allowed to do other activities when eating their meals

Source (health condition)	Purpose	Intervention description
Rahman et al, 2015 (Eating monitoring)	To detect a person's eating and drinking habits	Records a person's eating and drinking habits through head movements
		Helps people with obesity and diabetes
		Developed the Glass Eating and Motion (GLEAM) dataset
		Participants ate, walked, and did other activities during the monitoring period
		Participants did not interact with GG but simply wore it
		GG sensors recorded movement
Ye et al, 2015 (Eating monitoring)	To detail eating habits to help weight reduction	Collects images of the person's day from their perspective every 30 seconds
		Amazon's Mechanical Turk is a human computation platform that can determine eating behaviors and is used to identify when a person is eating
Hernandez et al, 2014 (Physiological measurements)	To measure heart rate and breaths per minute	Participant would wear GG, and GG's accelerometer, gyroscope, and camera were used to find user's pulse and respiratory rates
		The recording was done in several different positions including, sitting, standing, and lying down
Richer et al, 2015 (Physiological measurements)	To use the DailyHeart app to monitor ECGs	GG presents ECG signals to user in everyday life
		Signals are processed in real-time and classify the user's heart beats
		It will store data in an internal database
Wiesner et al, 2015 (Allergies)	To give consumers information of possible allergens in cosmetic products	An app is developed for GG whose purpose is to scan products
		User scans the product in the store and the GG app identifies the product
		User has uploaded the information of their specific allergies and the app compares the ingredients to the user's profile
		GG indicates whether the user should buy the product and why

Description of Google Glass Use as Clinician-Centered Interventions

Table 4 summarizes the Google Glass approach as clinician-centered interventions. Four of the clinician-focused studies (4/30, 13%) used Google Glass to assist in disaster relief by providing first responders with maps and navigational assistance [30], maintaining audiovisual communication with groups of paramedics and administrators [27], scanning triage tags [28], and performing teleconsultations with physician experts to confirm triage decisions [29]. Two studies (2/30, 7%) used Google Glass to help community health workers make more efficient diagnoses [32] and by allowing clinicians to retrieve images of similar cases [42]. Another (1/30, 3%) provided nurses information about peripheral stimuli to help them more efficiently manage their clinical environment [33]. Two studies (2/30, 7%) used the teleconsultation capabilities of Google Glass to improve the accuracy of neurological [34] and emergency dermatology [56] examinations. Nine studies (9/30, 30%) used Google Glass in student training situations to provide first-person demonstrations of procedures [37], record students in simulated patient interactions [38,49,52], enhance simulated interactions by projecting videos of the scenarios into their visual field [35,48], provide students with live instruction from an expert [40,46,55], and teach anatomy by providing real-time ultrasound imaging [47]. One study (1/30, 3%) used Google Glass to provide patient monitoring data to assist anesthesiologists and minimize distractions during procedures [36]. Five studies (5/30, 17%) used Google Glass to capture images of X-rays [39,44] and ECGs [43,45] that physicians then interpreted for significant findings. One study (1/30, 3%) used Google Glass to minimize head movements during ultrasound-guided procedures by projecting the images onto Google Glass [41]. One study (1/30, 3%) used Google Glass to take and analyze rapid diagnostic tests (RDTs) [50]. One study (1/30, 3%) helped clinicians evaluate burn patients by assisting in airway assessment [51]. One study (1/30, 3%) assessed the potential uses of Google Glass in autopsy or forensics settings by specifically evaluating the quality of images taken by Google Glass for documentation [53]. One study (1/30, 3%) leveraged multiple functions of Google Glass to assist with the treatment of chronic wounds [54]. Finally, one study (1/30, 3%) developed an app for Google Glass to facilitate behavioral, cognitive, and social research [31].

Table 4. Summary of Google Glass approach as clinician-centered interventions.

Source (health condition)	Purpose	Intervention description
Gillis et al, 2015 (Disaster relief)	To provide a hands-free way for doctors to be updated on the status and needed-care levels of critical-care patients	Developed a mesh network that covered a set area to allow communication between users and the hospital
		Users wore GG and could communicate with each other across the lake
		Users were then able to use the information they were getting in the field, record it, and relay it back to the hospital
Carenzo et al, 2014 (Disaster relief)	To aid in nontechnical skills in the management of disasters and mass casualty incidents	Used an app to GG to guide a Simple Triage and Rapid Treatment Triage visually
		Focused heavily on casualty identification, therefore the facial recognition capabilities for GG were used
		Visual information was then relayed to a secondary location for others to monitor
Cicero et al, 2014 (Disaster relief)	To streamline the triage system and then also offer consultations from an expert physician to those onsite	Paramedics used GG to communicate with an offsite physician disaster expert
		They assigned triage levels to victims using the SMART Triage System
		Offsite physician had an audio-video interface with paramedics so they could be observed in the offsite location
Newaz and Eide, 2015 (Disaster relief)	To provide direction to first responders in a new area	One group used GG as a tool for navigation
		The other group used a different device to navigate an unfamiliar neighborhood
		The route was preset on GG or the other device
Paxton et al, 2015 (Behavioral sciences)	To determine how interpersonal dynamics in conversation are affected by the environment	The app PsyGlass was created for GG
		The students wore GG and were presented with a series of red or blue lights as well as audio stimuli
		They had a conversation with the experimenter and their head movements were recorded through the GG accelerometer
Pappachan et al, 2014 (Diagnostics)	To assist community health workers to more efficiently diagnose patients	Uses Rafiki, a GG software that calculates age and gender and other characteristics to diagnose a patient
		Correlates between diseases, symptoms, and patients to determine the problem
Pascale et al, 2015 (Nursing – peripheral detection)	To help clinicians, such as nurses, pay attention to multiple patients while away from their station	Provided stimuli in the periphery of the nurses
		GG was used to detect and notify the nurses when something was presented in their peripheral vision
Yuan et al, 2015 (Neurology)	To make a neurological examination as accurate as possible through collaboration	A woman that suffered a right-sided dysphagia and asthenia was in the emergency department with a suspected stroke
		A local physician lacking neurological knowledge used GG to establish a teleconsult with a remote specialist who guided the physician in evaluating the patient
Chaballout et al, 2016 (Student training – health science students)	To teach health care students to respond to respiratory distress	Students watched a video while wearing GG
		Video showed a patient in respiratory distress
		Students then performed a procedure to aid respiratory distress on a manikin in front of them
Drake-Brockman et al, 2016 (Anesthesiology)	To allow anesthesiologists to monitor vitals of patients during procedures	AnaeVis was developed to run on GG, which provides visualization of patient monitoring data
		Anesthetists wore the device while treating the patient and the signals were shown and recorded

Source (health condition)	Purpose	Intervention description
Iversen et al, 2015 (Student training – physiotherapy students)	To record 1st-person view of procedures demonstrated by instructors to relay to students for training purposes	Faculty member wore GG during the performance of clinical skills
		Video of clinical skill performance was then shown to students for the purpose of teaching
Son et al, 2015 (Student training – otolaryngology residents)	To improve otolaryngology resident training by capturing 1st-person recordings of clinic encounters for later evaluation	Residents were recorded in an outpatient clinic by patients
		Patients were then given a survey to complete that rated their satisfaction level with their visit
		Video information was evaluated by two different parties and a review was given back to residents
Spaedy et al, 2016 (Radiology)	To improve the efficiency of remote chest X-ray interpretation	Fellows reviewed 12 chest X-rays with 23 major findings by viewing the image on GG, viewing an image taken by GG on a mobile device, and viewing the original X-ray on a desktop computer
		One point was given for each major finding
Russel et al, 2014 (Student training – medical students [radiology])	To determine if GG could provide telementoring instruction in bedside ultrasonography	Students wore GG and received real-time telementoring education
		Telementoring was done by an expert at a different location
		Students' goal was to obtain best parasternal long axis cardiac imaging using a portable GE Vscan
Wu et al, 2014 (Student training – medical students and radiology residents)	To minimize the amount of distraction caused by monitors during ultrasounds	Medical practitioner wore the GG during the ultrasound procedure
		GG screen projected images and video to the wearer
		Practitioner's hand movements and eye movement were recorded to see if there was improvement
Widmer et al, 2014 (Dermatology and Radiology)	To improve diagnostics in dermatology and cardiology	Participants would wear GG during a consultation
		ParaDISE app was developed to be a medical image retrieval system
		GG's visual and photo taking capabilities were utilized and then the photograph was sent into the interface and could be matched with similar images
		Those similar images were then sent to the wearer
Stetler et al, 2015 (Cardiology)	To capture and facilitate the interpretation of ECGs	ECGs were selected that had important findings
		GG zoom capabilities were used to identify each finding
		Every time a participant identified a finding they received one point
		ECGs were captured using the video function of GG
Duong et al, 2015 (Cardiology)	To facilitate the interpretation of coronary angiograms	GG's video function was used to record angiograms with specific findings
		Students were then told to try to determine each of the findings in the angiograms
Jeroudi et al, 2014 (Cardiology)	To facilitate the interpretation of ECGs	Physicians wore GG and looked at the ECG image on the screen
		Physicians wore GG and viewed a photograph of the ECG taken using GG and then viewed on a mobile device
		Results were then compared to other methods of viewing ECGs
Vallurupalli et al, 2013, (Student training – medical students [cardiology])	To improve resident training by streaming the view of residents during simulations to attending physicians for consultation	Residents wore GG while working through four scenarios in cardiovascular practice
		Live video of the scenarios taken by GG was streamed to a mobile phone or personal computer used by the attending physician

Source (health condition)	Purpose	Intervention description
Benninger, 2015 (Radiology)	To facilitate teaching anatomy to medical students	Students familiarized themselves with GG for 10-30 minutes using a program called MiniGames
		Students were then given tutorials in groups of 3-5 while using GG with a finger probe to identify neuromuscular and organ structures and spaces in the limbs and cavities
		Students were tested during 7 separate laboratory examinations over 1 year to identify the same structures and practice procedures
Vaughn et al, 2016 (Student training – nursing students)	To increase the perception of realism in nursing student simulations	Students were allowed 10 minutes to familiarize themselves with GG before the intervention
		Students were then given the patient report and started the simulation in which GG projected a video of an acute asthma exacerbation scenario
		1-2 Certified Healthcare Simulation Experts evaluated students' performance
Zahl et al, 2016 (Student training – dental students)	To facilitate self- and peer-assessment of standardized patient (SP) interactions for dental students	3rd-year dental students volunteered to record their SP encounter using GG while a traditional static camera simultaneously recorded
		All GG and static camera videos were later reviewed during Behavioral Patient Management small group discussions
		Students rated how effective each type of video was for assessing communication skills
Feng et al, 2015 (Diagnostics – HIV or cancer)	To improve the efficiency of immunochromatographic diagnostic test analysis	One or more RDTs, either HIV (qualitative) or PSA (quantitative), labeled with QR codes were imaged using GG
		Images were automatically transmitted to a digital server that located all RDTs and produced a quantitative diagnostic result, which was reported to user
Spencer et al, 2014 (Pulmonology – airway assessment for burn victims)	To facilitate airway assessment of burn patients requiring surgery	GG was worn by physicians during two cases of burn patients requiring airway assessment
		Documentation of procedure by GG was evaluated after the intervention
Tully et al, 2015 (Student training – medical students [hospice])	To facilitate medical student self-evaluation after end-of-life SP encounters	2nd-year medical students participated in end-of-life SP encounters where the SP was wearing GG to record the encounter
		Students then reviewed GG and traditional videos
Albrecht et al, 2014 (Pathology – autopsy and postmortem examinations)	To evaluate the feasibility of using GG in a forensics setting	Two physicians wore GG during 4 autopsy and postmortem examinations and took images using both GG and a traditional digital single lens reflex (DSLR) camera
		Six forensic examiners evaluated the images for quality
Aldaz et al, 2015 (Chronic wounds)	To facilitate photo documentation of chronic wounds for long-term care	Wound care nurses used SnapCap software on GG to take images, tag, and transfer them to patient electronic medical records
		Image quality and ease of use were evaluated
Chai et al, 2015 (Toxicology)	To facilitate toxicology teleconsultation in the emergency department	Emergency medicine residents wore GG while evaluating poisoned patients
		Real-time video of physician findings was transmitted to toxicology fellows and attendings for evaluation
Chai et al, 2014 (Dermatology)	To facilitate dermatology teleconsultation in the emergency department	Patients first had a standard dermatology consultation (phone call and sometimes a static photo of the rash) with a dermatology resident
		Patients were then evaluated by the dermatology chief resident through a real-time video filmed by the patient (wearing GG) and viewed by the physician on a tablet

Feasibility and Acceptability of Google Glass as Patient-Centered Interventions

Table 5 summarizes the user satisfaction results of the patient-centered interventions (see Multimedia Appendix 3 for more technical results). Overall, participant feedback on the comfort and ease of use of Google Glass in patient-centered interventions was very positive. Of the participants with visual impairments, Anam et al reported a median usability score of 4.6/5 [14], and Tanuwidjaja reported that a majority of participants believed the Google Glass intervention would be useful in everyday life [18]. Among the participants with motor impairments, namely Parkinson's Disease (PD), while overall reactions were positive [12] and some believed that Google Glass allowed them to do things they were not previously able to do independently [11], there were also some consistent frustrations expressed. For example, some experienced difficulties using the touchpad and voice navigation features as a result of tremors and dysarthria associated with the disease [10,11]. While participants with ASD reported positive experiences using Google Glass [20], participants using it to collect physiological data reported privacy concerns and found a smartwatch to have better usability [25]. Some of the common complaints reported were overheating of the device [14,15,20], its relatively short battery life [6,14,15,17,18,22,24], poor quality camera [17,18], perceived stigma when wearing the device in public [14,25], wireless connectivity issues [26], and concerns about privacy and the protection of confidential information [11,25]. These results suggest the potential for future research and implementation to support patients if Google can address some of the device's technological limitations. However, the issues experienced by PD patients due to dysarthria and tremors should be addressed in interventions related to motor impairments.

Table 5. Feasibility and acceptability of Google Glass as patient-centered interventions.

Source (health condition)	User satisfaction results
Anam et al, 2014 (Ophthalmology – visual impairment)	Participants completed 5-point Likert scale on usability of the Expression system (a score of 5=the best): Learnability median 4.1, interquartile range (IQR) 0.7; Informativeness median 4.5, IQR 1.0; Usability median 4.6, IQR 0.7; User Satisfaction median 4.5, IQR 1.0; Willing to Use median 3.7, IQR 0.7. The relatively low score for "Willing to Use" can be attributed to perceived uncertainty in social acceptability of wearing a device such as GG.
Tanuwidjaja et al, 2014 (Ophthalmology – colorblindness)	4/6 participants reported they found Chroma system useful in performing study tasks and would find it useful in everyday life. Two participants expressed concerns about system lag time in switching between modes. One participant did not find the system helpful because his vision test scores worsened when using Chroma.
Malu and Findlater, 2015 (Motor impairment – upper body)	Participants rated system features on a 5-point scale (1=very easy/comfortable to 5=very difficult/uncomfortable): Visual Display comfort median 2, mean 2.2, SD 1.2; ease median 2, mean 2.2, SD 1.2; Touchpad Gestures comfort median 3, mean 3, SD 2.2; ease median 2, mean 2.7, SD 1.9; Voice Commands ease median 1, mean 1.7, SD 1.2. For the reciprocal tapping task, most (N=8) found the large touchpad easiest to use, and most (N=7) found the large touchpad to be most physically comfortable.
McNaney et al, 2014 (Motor impairment – PD)	Study exit interviews identified some concerns with usability of and patient satisfaction with GG. Some felt wearing GG in public drew unwanted attention, and 3/4 participants reported they would not wear GG in certain settings due to safety concerns. All participants experienced frustration when certain features, such as voice recognition and navigation, were difficult to use in everyday life or did not work. However, when the features were working properly, user satisfaction was high. GG enabled some to do things others without PD can do on mobile phones. Overall, reactions to GG were positive and showed appreciation for how GG could be used to help those with PD.
McNaney et al, 2015 (Motor impairment – PD)	Study exit interviews revealed mixed reactions to LApp program, with some finding significant improvement in and confidence with their speech volume and others reporting the program performance was inconsistent. Additional frustrations were related to GG's short battery life and difficulties navigating the touchpad because of PD-related tremors.
Zhao et al, 2016 (Motor impairment – PD)	Most users found GG easy or very easy to use (N=7/11) and the instructions on screen clear or very clear to read (9/12). One user particularly liked the bone-conducting headphone because the metronome was less audible to others around. Some participants disliked GG's placement of the visual display in the upper right corner (n=3) and suggested images be projected binocularly (n=1) or more focally (n=2) in the visual field. They suggested verbal instructions (n=9), rhythmic music (n=2), and postural feedback (n=1) as additional cues for the app and that cues be provided only when needed (n=2).
Voss et al, 2016 (Psychological/Developmental – ASD)	Review of videos of participants using GG system at home showed that children reported positive experiences with the activities at home and stated they viewed the system as a toy. However, the device heated up to uncomfortable levels if worn too long.
Richer et al, 2015 (Physiological measurements)	Participants completed a qualitative assessment of their experience using *DailyHeart* on GG. Mean usability rating for smartwatches (4.2) was higher than GG rating (2.8). Almost a third of participants were afraid that health data stored in Google Fit could be misused by third parties. In comparing the use of *DailyHeart* on GG and on Android Wear, Wear outperformed GG on all measures (appearance, features, handling, distraction, and overall usability).

Feasibility and Acceptability of Google Glass as Clinician-Centered Interventions

Table 6 summarizes user satisfaction and technical results of the clinician-centered interventions (see Multimedia Appendix 4 for more technical results). The clinician-centered studies, which also varied greatly in specializations and uses of Google Glass, reported more inconsistent reactions regarding the utility of the device. Many of the studies reported technical frustrations similar to those mentioned in the patient-centered interventions—specifically, short battery life [29,31,35,37,41,42,48,50,53], device overheating [31,35,37,41], difficulties with wireless network connection [29,35,37,41,42,48,55], and privacy concerns [28,31]. Multiple studies reported that, while the device was generally comfortable to wear and did not distract from the clinician's work, it did not significantly improve outcomes or the clinician's efficiency [29,37,53]. The main sources of frustration specific to clinician use were the size and quality of images taken with and viewed through the device [42,45,50], difficulties in taking images and videos, and keeping patient monitors in the clinician's line of vision due to the fact that the Google Glass follows a person's head movements instead of gaze [33,36,49]. Specifically, Spaedy et al found that clinicians were dissatisfied viewing images of chest X-rays through Google Glass but were impressed with the images taken by the device and viewed on a mobile phone or computer [39]. In addition, Stetler et al, Duong et al, and Jeroudi et al reported that cardiologists were generally not confident with their interpretations of ECGs viewed through Google Glass [43-45]. However, other studies, in particular the ones that used Google Glass as a tool for training students, found that students had overall positive reactions to Google Glass and would recommend its future use [35,36,41,47-49,52]. For example, Chaballout et al found that most students (10/12, 83%) recommended its continued use in clinical simulations [35]. Similarly, Wu et al found that a majority (88%) of the medical students and radiology residents would be likely to use ultrasound visualization through Glass instead of a traditional monitor [41]. One exception to this trend was a study by Iversen et al, which reported that a majority of the physiotherapy students (26/39, 67%) found Google Glass' video quality unacceptable, and many (23/39, 59%) did not feel the device enhanced their learning experience [37]. Despite this contradiction, these results suggest that the greatest potential for Google Glass implementation to support clinicians lies in student training.

Table 6. Feasibility and acceptability of Google Glass as clinician-centered interventions.

Source (health condition)	User satisfaction results
Cicero et al, 2014 (Disaster relief)	First responders using GG completed a survey assessment after the intervention, and their responses supported the idea that GG does not make a significant improvement in disaster triage.
Yuan et al, 2015 (Neurology)	Local physicians found that holding a mobile phone to provide the consulting specialist live images on GG was inconvenient.
	Teleneurohospitalists using GG did not feel the system allows for patient evaluation similar to what would be achieved in-person.
Chaballout et al, 2016 (Student training – health science students)	Participants were asked to complete 2 post-intervention surveys, a 13-item Student Satisfaction and Self-Confidence in Learning Scale and a 20-item Simulation Design Scale (scale for both measures was 1=strongly disagree to 5=strongly agree). Most students recommended continued use of GG in clinical simulations (N=10/12). They also reported high mean scores on the simulations's design and satisfaction with the simulation to promote learning and self-confidence in learning.
	Simulation Design Scale (mean [SD]): Objectives and information 4.65 (0.18); Support 4.85 (0.04); Problem solving 4.53 (0.30); Feedback/guided reflection 4.85 (0.14); Fidelity (realism) 4.67 (0.12)
	Student Satisfaction and Self-Confidence with Learning (mean [SD]): Satisfaction with current learning 4.67 (0.13); Self-confidence in learning 4.35 (0.60).
Drake-Brockman et al, 2016 (Anesthesiology)	Anesthetists participating in the intervention were asked to complete a survey including a Likert scale and freeform questions: 78% would use GG again, 58% would recommend GG to colleagues, 21% felt GG improved patient management, 90% reported GG was comfortable to wear, 86% reported that information presented on GG was easy to read, 56% would wear GG in view of patients, 75% felt positive about using GG in the operating room environment, 82.5% reported that wearing GG did not distract from patient management.
Iversen et al, 2015 (Student training – physiotherapy students)	Students who used GG in the study answered questions about the technology after the intervention. 67% (26/39) of students evaluated GG video quality as not acceptable (score of ≤2 on the Likert scale), and 59% (23/39) of students reported using GG did not enhance their learning experience.
Spaedy et al, 2016 (Radiology)	Participants responded to a 5-point Likert scale about the quality of GG images and their confidence about their interpretation. When viewing images through GG, 87% (13/15) were dissatisfied with the image and unsure that such a small display would be able to provide the necessary level of detail. 80% (12/15) were impressed with image clarity taken via GG and viewed on the mobile device.
Wu et al, 2014 (Student training – medical students and radiology residents)	Participants who used GG responded to a post-exercise survey. 87% reported GG was comfortable to use for ultrasound guidance. 88% reported they would be likely to use ultrasound visualization through GG as opposed to traditional monitors (18% very likely, 35% moderately likely, 35% somewhat likely). 78% indicated they would "very likely" be interested in future research studies involving GG in medical simulation and education.
Stetler et al, 2015 (Cardiology)	Physicians responded to a 5-point user-experience Likert scale after the intervention. 58% (7/12) were satisfied with GG image quality of ECGs. 50% (6/12) were confident in their interpretation when using GG.
Duong et al, 2015 (Cardiology)	Participants responded to a post-study survey regarding their satisfaction with image quality and comfort making clinical recommendations. 10% (1/10) were "neutral" regarding quality and giving recommendations. 60% (6/10) of physicians were "somewhat satisfied" and would be "somewhat comfortable" giving recommendations. 30% (3/10) were "very satisfied" and would be "very comfortable" giving recommendations.
Jeroudi et al, 2014 (Cardiology)	Participants completed subjective ratings on a 5-point Likert scale regarding image quality and their confidence of ECG interpretation. 75% (9/12) were dissatisfied with the ECG image quality when viewing via GG. 83% (10/12) were not confident in their interpretation when viewing via GG. 58% (7/12) were neutral about ECG images taken by GG and viewed on mobile phones. 58% (7/12) were more confident in their interpretation when viewing the GG image on a mobile phone than when viewing via GG.
Benninger, 2015 (Radiology)	Participants responded to a 5-point Likert scale questionnaire. Did they enjoy the exposure to technology applying the triple feedback method? Average score 4.6. Would they prefer more time with the technology? Average score 4.8
Vaughn et al, 2016 (Student training – nursing students)	After the intervention, students responded to 2 surveys, the Simulation Design Scale and the Self-Confidence in Learning Scale (both 5-point scales from 1=strongly disagree to 5=strongly agree), to assess their perception of GG in the simulation: Independent problem-solving was facilitated, 4.75 (0.45); Resembled a real-life situation, 4.75 (0.45); Teaching methods were helpful and effective, 4.67 (0.65); Teaching materials were motivating and helpful, 4.58 (0.90); Confidence in mastering simulation content: 4.42 (0.51); Develops skills/knowledge applicable to a clinical setting, 4.83 (0.39)

Source (health condition)	User satisfaction results
Zahl et al, 2016 (Student training – dental students)	Students responded to 4 open- and closed-text items about using GG and static video for self- and peer-assessment. Students' reported mean score was higher for GG recordings (84.61) than static video (79.74). Students reported that verbal communication was more easily assessed by reviewing GG video (23.87) than static video (22.17); paraverbal communication was more easily assessed by reviewing GG video (24.26) than static video (21.51); and nonverbal communication was more easily assessed by reviewing static video (19.78) than GG video (17.09).
Tully et al, 2015 (Student training – medical students [hospice])	Students responded to a 5-point Likert scale on how distracting they found GG during the intervention. 23% (7/30) reported a "positive, nondistracting experience." 37% (11/30) reported a "positive, initially distracting experience." 17% (5/30) reported a "neutral experience." 10% (3/30) reported a "negative experience." After reviewing the videos filmed with GG, 70% (16/30) believed that GG is worth including in the clinical skills training program.
Albrecht et al, 2014 (Pathology – autopsy and postmortem examinations)	Both participants agreed that GG was comfortable to wear but required more physical effort to capture images than a DSLR camera.
Chai et al, 2015 (Toxicology)	Study participants completed a survey immediately after the consult about their experience viewing a teleconsult through GG: 94% (17/18) were confident in the toxidrome after GG consultation as compared to 56% (10/18) who were confident after phone consultation.
Chai et al, 2014 (Dermatology)	All participants responded to a survey on acceptability of GG after their consultation. 93.5% (29/31) were overall satisfied with the video consultation. 22.6% (7/31) preferred care provided through mobile video communication technology over a standard face-to-face clinic visit. 74.2% (23/31) preferred care provided through mobile video communication technology over standard emergency department telephone consultation. 93.3% (28/31) would recommend the video consultation to others. 96.8% (28/30) felt comfortable that privacy was protected during the video encounter. 96.8% (30/31) were confident in the video equipment used.

Discussion

Principal Findings

In recent years, wearable devices such as wrist-worn accelerometers and head-mounted devices have become increasingly popular for their applications to everyday life as well as to various industries. While Google Glass, one of the more well-known head-mounted wearable devices, has yet to successfully break into the consumer market, various industries are eager to harness its potential in their fields. Medicine is one such industry; however, far greater attention has been paid to surgical applications than to nonsurgical ones. In this systematic review, we assessed existing evidence of the usability, benefits, and limitations of Google Glass to support both patients and clinicians in nonsurgical medical settings. Overall, the evidence was somewhat limited by a small number of studies fitting all inclusion criteria, small sample sizes, and other methodological considerations, particularly for statistical analysis. We included 51 studies that met our pre-set inclusion criteria, with the majority of studies describing clinician-centered interventions. There was a wide range of health conditions and uses of Google Glass. While information regarding age of participants was limited, the studies that did include age information were conducted with adults and none within pediatric populations. Many were conducted in laboratory, hospital, and student training settings, which indicates potential of university-affiliated teaching hospitals to integrate wearable technologies to make clinicians more efficient and provide clinical support to patients.

Unlike our systematic review, other recent reviews of the use of wearable technology in medicine included other heads-up devices besides Google Glass and did not distinguish between surgical and nonsurgical interventions [3,4]. Some of the uses of Google Glass in these studies include data visualization and video recording during surgery and interventional radiology, smart checklists, telementoring, virtual reality for education or pain management, interpretation of images, teleconsultation, teleconferencing, drug delivery tracking, patient empowerment, laboratory diagnostics, and forensic medicine [3,4].

A recent systematic review of medical applications of Google Glass in both surgical and nonsurgical settings found more globally positive support for the technology's use in these settings [4]. However, this systematic review discussed a smaller sample of articles (n=21) that spanned surgical and nonsurgical medical interventions as well as scientific settings in general. Furthermore, original research studies on Google Glass in surgical interventions report fewer technical issues with the device and recommend strategies to overcome those that were encountered. One study, in which a pediatric surgeon wore Google Glass continuously for 4 weeks, reports with confidence that the photographic and video quality of the device was sufficient to capture all clinically relevant findings [57]. In contrast, many of the studies included in our analysis cited the photographic and video quality of Glass as a significant clinical limitation. While our findings regarding the limitations to the use of Glass, namely battery life, photographic and video quality, and streaming capabilities, were consistent with those encountered in surgical applications, the surgical studies seem to have made more progress in testing potential solutions. For example, one plastic surgeon used a USB-powered pocket battery to eliminate the need to recharge the device during the operation, noise-canceling headphones to enhance the sound transmitted by Glass, and a light emitting diode (LED) lamp headset to improve photo and video quality [4]. These findings support the potential of Google Glass to be even more beneficial

in nonsurgical medical interventions if technical limitations are overcome either in newer models of the device or with the implementation of these solutions.

Strengths and Limitations

Our systematic review has a number of strengths. First, our review was conducted following the recommendations and guidelines for rigorous systematic reviews methodology [58-60]. Second, we used a very sensitive search strategy guided by a librarian information specialist with no language restrictions to include as many relevant studies as possible and minimize possible publication bias. In addition, we searched other resources, including published systematic reviews, clinical trial registries, and different electronic databases. Finally, 2 authors completed the review process independently at all stages.

Our systematic review of the literature has some potential methodological limitations. First, similar to other systematic reviews, although our search criteria were comprehensive, we could have missed some relevant articles [61]. Second, we included only original research papers that have been published in peer-reviewed journals, and the possibility of publication bias with the tendency to report positive study results cannot be excluded [62]. Finally, a number of the studies included in our review had a relatively small sample size.

Conclusions

Results regarding the feasibility, usability, and acceptability of Google Glass in nonsurgical medical settings were extremely varied, with more positive results being reported for patient-centered studies and student training settings. Further investigation with rigorous research designs evaluating the efficacy and cost-effectiveness of these more successful interventions in supporting patients and clinicians is warranted. These efforts would be beneficial in informing the base of evidence on the use of wearable devices, such as Google Glass, in medicine.

Acknowledgments

We thank Ms Linda O'Dwyer (Galter Health Sciences Library, Northwestern University Feinberg School of Medicine, Chicago, IL) for her support with the literature search. We also thank Ms Ayana Ceaser (Northwestern University Weinberg College of Arts and Sciences, Evanston, IL) for her help with data extraction.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA checklist.

[PDF File (Adobe PDF File), 65KB - mhealth_v5i10e159_app1.pdf]

Multimedia Appendix 2

Search strategies.

[PDF File (Adobe PDF File), 27KB - mhealth_v5i10e159_app2.pdf]

Multimedia Appendix 3

Summary of the technical results of the patient-centered studies.

[PDF File (Adobe PDF File), 80KB - mhealth_v5i10e159_app3.pdf]

Multimedia Appendix 4

Summary of the technical results of the clinician-centered studies.

[PDF File (Adobe PDF File), 94KB - mhealth_v5i10e159_app4.pdf]

References

1. Iqbal MH, Aydin A, Brunckhorst O, Dasgupta P, Ahmed K. A review of wearable technology in medicine. J R Soc Med 2016 Oct;109(10):372-380. [doi: 10.1177/0141076816663560] [Medline: 27729595]
2. Michael A, Tehrani K. Wearable Technology and Wearable Devices: Everything You Need to Know. 2014. URL: http://www.wearabledevices.com/what-is-a-wearable-device/ [accessed 2017-08-27] [WebCite Cache ID 6t1sQoa7o]
3. Wrzesinska N. MEDtube Science. The use of smart glasses in healthcare URL: https://medtube.net/science/wp-content/uploads/2014/03/The-use-of-smart-glasses-in-healthcare-%E2%80%93-review.pdf [accessed 2017-08-27] [WebCite Cache ID 6t1sqqx1p]

4. Davis CR, Rosenfield LK. Looking at plastic surgery through Google Glass: part 1. Systematic review of Google Glass evidence and the first plastic surgical procedures. Plast Reconstr Surg 2015 Mar;135(3):918-928. [doi: 10.1097/PRS.0000000000001056] [Medline: 25719707]
5. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
6. Lazewatsky D, Bowie C, LaFortune J. Wearable computing to enable robot microinteractions. 2014 Presented at: International Symposium on Robot and Human Interactive Communication; Aug 25-29, 2014; Edinburgh, UK p. 141-146 URL: http://ieeexplore.ieee.org:80/document/6926244/?reload=true
7. Gips J, Zhang M, Anderson D. Towards a Google Glass Based Head Control Communication System for People with Disabilities. In: Intephanidis C, editor. Communications in Computer and Information Science. Switzerland: Springer International Publishing; Jul 12, 2015:978-983.
8. Sinyukov D, NW RT. Augmenting a voice and facial expression control of a robotic wheelchair with assistive navigation. 2014 Presented at: International Conference on Systems, Man and Cybernetics (SMC); Oct 5-8, 2014; San Diego, CA p. 1088-1094 URL: https://doi.org/10.1109/SMC.2014.6974059
9. Malu M, Findlater L. Personalized, Wearable Control of a Head-mounted Display for Users with Upper Body Motor Impairments. 2015 Presented at: Annual ACM Conference on Human Factors in Computing Systems; Apr 18-23, 2015; Seoul, Republic of Korea p. 221-230 URL: https://doi.org/10.1145/2702123.2702188 [doi: 10.1145/2702123.2702188]
10. McNaney R, Poliakov I, Vines J, Balaam M, Zhang P, Olivier P. LApp: A Speech Loudness Application for People with Parkinson's on Google Glass. 2015 Presented at: Annual ACM Conference on Human Factors in Computing Systems; Apr 18-23, 2015; Seoul, Republic of Korea p. 497-500 URL: https://doi.org/10.1145/2702123.2702292
11. McNaney R, Vines J, Roggen D, Balaam M, Zhang P, Poliakov I, et al. Exploring the acceptability of google glass as an everyday assistive device for people with parkinson's. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; Apr. 26-May 1, 2014; Toronto, Ontario, Canada p. 2551-2554 URL: https://dl.acm.org/citation.cfm?id=2557092 [doi: 10.1145/2556288.2557152]
12. Zhao Y, Nonnekes J, Storcken EJM, Janssen S, van WEEH, 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]
13. Pervaiz M, Patel R. SpeechOmeter: heads-up monitoring to improve speech clarity. 2014 Presented at: International ACM SIGACCESS Conference on Computers & Accessibility; Oct 20-22, 2014; Rochester, New York p. 319-320 URL: https://doi.org/10.1145/2661334.2661339
14. Anam A, Alam S, Yeasin M. Expression: A Dyadic Conversation Aid using Google Glass for People with Visual Impairments. 2014 Presented at: International Conference on Mobile Computing, Applications and Services (MobiCASE); Nov. 6-7, 2014; Austin, TX p. 211-214 URL: https://doi.org/10.4108/icst.mobicase.2014.257780
15. Garcia G, Nahapetian A. Wearable computing for image-based indoor navigation of the visually impaired. In: Wireless Health Conference. 2015 Presented at: Wireless Health; Oct 14-16, 2015; Bethesda, MD p. 6. [doi: 10.1145/2811780.2811959]
16. Pundlik S, HuaQi Y, Rui L, Peli E, Gang L. Magnifying Smartphone Screen Using Google Glass for Low-Vision Users. IEEE Trans Neural Syst Rehabil Eng 2017 Jan;25(1):52-61. [doi: 10.1109/TNSRE.2016.2546062] [Medline: 28113862]
17. Hwang AD, Peli E. An augmented-reality edge enhancement application for Google Glass. Optom Vis Sci 2014 Aug;91(8):1021-1030 [FREE Full text] [doi: 10.1097/OPX.0000000000000326] [Medline: 24978871]
18. Tanuwidjaja ED, Koa KC, Shao CPCN. Chroma: A Wearable Augmented-Reality Solution for Color-Blindness. 2014 Presented at: UbiComp; Sept. 13-17, 2017; Seattle, WA p. 799-810 URL: http://ubicomp.org/ubicomp2014/ [doi: 10.1145/2632048.2632091]
19. Miranda D, Calderón M, Favela J. Anxiety detection using wearable monitoring. 2014 Presented at: MexIHC; Nov. 3-5, 2014; Oaxaca, Mexico p. 34-41 URL: http://www.mexihc.org/2014/EN [doi: 10.1145/2676690.2676694]
20. Voss C, Washington P, Haber N. Superpower glass: delivering unobtrusive real-time social cues in wearable systems. 2016 Presented at: UbiComp; Nov. 12-16, 2016; Heidelberg, Germany p. 1218-1226 URL: http://ubicomp.org/ubicomp2016/ [doi: 10.1145/2968219.2968310]
21. Mirtchouk M, Merck C, Kleinberg S. Automated estimation of food type and amount consumed from body-worn audio and motion sensors. 2016 Presented at: UbiComp; Sept. 12-16, 2016; Heidelberg, Germany p. 451-462 URL: http://ubicomp.org/ubicomp2016/ [doi: 10.1145/2494091.2494143]
22. Rahman S, Merk C, Huang Y, Kleinberg S. Unintrusive Eating Recognition using Google Glass. 2015 Presented at: PervasiveHealth; May 20-23, 2015; Istanbul, Turkey p. 108-111 URL: http://pervasivehealth.org/2015/show/home [doi: 10.4108/icst.pervasivehealth.2014.255423]
23. Ye X, Chen G, Cao Y. Automatic Eating Detection using head-mount and wrist-worn accelerometers. : IEEE; 2015 Presented at: International Conference on E-health Networking, Application & Services (HealthCom); Oct. 14-17, 2015; Boston, MA p. 578-581. [doi: 10.1109/HealthCom.2015.7454568]

24. Hernandez J, Li Y, Rehag JM, Picard RW. BioGlass: Physiological Parameter Estimation Using a Head-mounted Wearable Device. EAI International Conference on. 2014 Presented at: International Conference on Wireless Mobile Communication and Healthcare (Mobihealth); Nov. 3-5, 2014; Athens, Greece URL: https://doi.org/10.1109/MOBIHEALTH.2014.7015908
25. Richer R, Maiwald T, Pasluosta C, Hensel B, Eskofier B. Novel human computer interaction principles for cardiac feedback using google glass and Android wear. 2015 Presented at: International Conference on Wearable and Implantable Body Sensor Networks (BSN); June 9-12, 2015; Cambridge, MA p. 1-6. [doi: 10.1109/BSN.2015.7299363]
26. Wiesner M, Pobiruchin M, Hetterich C, Pfeifer D. GlassAllergy: a Google Glass-based solution to empower patients with skin allergies. Stud Health Technol Inform 2014;205:548-552. [Medline: 25160245]
27. Gillis J, Calyam P, Bartels A. Panacea's Glass: Mobile Cloud Framework for Communication in Mass Casualty Disaster Triage. 2015 Presented at: International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud); Mar. 30-Apr. 3, 2015; San Francisco, CA p. 128-134 URL: https://doi.org/10.1109/MobileCloud.2015.39
28. Carenzo L, Barra FL, Ingrassia PL, Colombo D, Costa A, Della CF. Disaster medicine through Google Glass. Eur J Emerg Med 2015 Jun;22(3):222-225. [doi: 10.1097/MEJ.0000000000000229] [Medline: 25460812]
29. Cicero MX, Walsh B, Solad Y, Whitfill T, Paesano G, Kim K, et al. Do you see what I see? Insights from using google glass for disaster telemedicine triage. Prehosp Disaster Med 2015 Feb;30(1):4-8. [doi: 10.1017/S1049023X1400140X] [Medline: 25571779]
30. Newaz F, Eide A, Pultier A. Supporting first responder in-field communication and navigation using head-mounted displays. 2015 Presented at: International Conference on Information Systems for Crisis Response and Management (ISCRAM); May 24-27, 2015; Kristiansand, Norway URL: http://idl.iscram.org/files/fahdbinmaleknewaz/2015/1208_FahdBinMalekNewaz_etal2015.pdf
31. Paxton A, Rodriguez K, Dale R. PsyGlass: Capitalizing on Google Glass for naturalistic data collection. Behav Res Methods 2015 Sep;47(3):608-619. [doi: 10.3758/s13428-015-0586-z] [Medline: 25893865]
32. Pappachan P, Yus R, Joshi A, Finin T. Rafiki: A Semantic and Collaborative Approach to Community Health-Care in Underserved Areas. 2014 Presented at: International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom); Oct. 22-25, 2014; Miami, FL URL: https://doi.org/10.4108/icst.collaboratecom.2014.257299 [doi: 10.4108/icst.collaboratecom.2014.257299]
33. Pascale M, Sanderson P, Liu D, Mohamed I, Stigter N, Loeb R. Peripheral Detection for Abrupt Onset Stimuli Presented via Head-Worn Display. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2015;59(1):1326-1330 [FREE Full text]
34. Yuan Z, Liu Z, Wei D, Shi M, Wang B, Liu Y, et al. Mobile Stroke: An Experience of Intravenous Thrombolysis Guided by Teleconsultation Based on Google Glass. CNS Neurosci Ther 2015 Jul;21(7):607-609. [doi: 10.1111/cns.12413] [Medline: 26096048]
35. Chaballout B, Molloy M, Vaughn J, Brisson IR, Shaw R. Feasibility of Augmented Reality in Clinical Simulations: Using Google Glass With Manikins. JMIR Med Educ 2016 Mar 07;2(1):e2 [FREE Full text] [doi: 10.2196/mededu.5159] [Medline: 27731862]
36. Drake-Brockman TFE, Datta A, von UBS. Patient monitoring with Google Glass: a pilot study of a novel monitoring technology. Paediatr Anaesth 2016 May;26(5):539-546. [doi: 10.1111/pan.12879] [Medline: 26992465]
37. Iversen M, Kiami S, Singh K, Masiello I, von Heideken HJ. Prospective, randomised controlled trial to evaluate the effect of smart glasses on vestibular examination skills. BMJ Innovations 2016;2:99-105 [FREE Full text] [doi: 10.1136/bmjinnov-2015-000094]
38. Son E, Halbert A, Abreu S, Hester R, Jefferson G, Jennings K, et al. Role of Google Glass in improving patient satisfaction for otolaryngology residents: a pilot study. Clin Otolaryngol 2017 Apr;42(2):433-438. [doi: 10.1111/coa.12810] [Medline: 27992944]
39. Spaedy E, Christakopoulos GE, Tarar MNJ, Christopoulos G, Rangan BV, Roesle M, et al. Accuracy of remote chest X-ray interpretation using Google Glass technology. Int J Cardiol 2016 Sep 15;219:38-40. [doi: 10.1016/j.ijcard.2016.05.070] [Medline: 27262231]
40. Russell PM, Mallin M, Youngquist ST, Cotton J, Aboul-Hosn N, Dawson M. First “glass” education: telementored cardiac ultrasonography using Google Glass - a pilot study. Acad Emerg Med 2014 Nov;21(11):1297-1299 [FREE Full text] [doi: 10.1111/acem.12504] [Medline: 25377415]
41. Wu TS, Dameff CJ, Tully JL. Ultrasound-guided central venous access using Google Glass. J Emerg Med 2014 Dec;47(6):668-675. [doi: 10.1016/j.jemermed.2014.07.045] [Medline: 25281180]
42. Widmer A, Schaer R, Markonis D, Muller H. Facilitating medical information search using Google Glass connected to a content-based medical image retrieval system. Conf Proc IEEE Eng Med Biol Soc 2014;2014:4507-4510. [doi: 10.1109/EMBC.2014.6944625] [Medline: 25570993]
43. Stetler J, Resendes E, Martinez-Parachini JR, Patel K, Amsavelu S, Tarar MNJ, et al. Hands-free zoom and pan technology improves the accuracy of remote electrocardiogram interpretation using Google Glass. Int J Cardiol 2016 Feb 01;204:147-148. [doi: 10.1016/j.ijcard.2015.11.144] [Medline: 26657610]

44. Duong T, Wosik J, Christakopoulos GE, Martínez Parachini JR, Karatasakis A, Tarar MNJ, et al. Interpretation of Coronary Angiograms Recorded Using Google Glass: A Comparative Analysis. J Invasive Cardiol 2015 Oct;27(10):443-446 [FREE Full text] [Medline: 26429845]
45. Jeroudi OM, Christakopoulos G, Christopoulos G, Kotsia A, Kypreos MA, Rangan BV, et al. Accuracy of remote electrocardiogram interpretation with the use of Google Glass technology. Am J Cardiol 2015 Feb 01;115(3):374-377. [doi: 10.1016/j.amjcard.2014.11.008] [Medline: 25482681]
46. Vallurupalli S, Paydak H, Agarwal SK, Agrawal M, Assad-Kottner C. Wearable technology to improve education and patient outcomes in a cardiology fellowship program - a feasibility study. Health Technol 2013 Oct 11;3(4):267-270. [doi: 10.1007/s12553-013-0065-4]
47. Benninger B. Google Glass, ultrasound and palpation: the anatomy teacher of the future? Clin Anat 2015 Mar;28(2):152-155. [doi: 10.1002/ca.22480] [Medline: 25377631]
48. Vaughn J, Lister M, Shaw RJ. Piloting Augmented Reality Technology to Enhance Realism in Clinical Simulation. Comput Inform Nurs 2016 Sep;34(9):402-405. [doi: 10.1097/CIN.0000000000000251] [Medline: 27258807]
49. Zahl DA, Schrader SM, Edwards PC. Student perspectives on using egocentric video recorded by smart glasses to assess communicative and clinical skills with standardised patients. Eur J Dent Educ 2016 Jul 06. [doi: 10.1111/eje.12217] [Medline: 27380732]
50. Feng S, Caire R, Cortazar B, Turan M, Wong A, Ozcan A. Immunochromatographic diagnostic test analysis using Google Glass. ACS Nano 2014 Mar 25;8(3):3069-3079 [FREE Full text] [doi: 10.1021/nn500614k] [Medline: 24571349]
51. Spencer RJ, Chang PH, Guimaraes AR, Firth PG. The use of Google Glass for airway assessment and management. Paediatr Anaesth 2014 Sep;24(9):1009-1011. [doi: 10.1111/pan.12496] [Medline: 25039494]
52. Tully J, Dameff C, Kaib S, Moffitt M. Recording medical students' encounters with standardized patients using Google Glass: providing end-of-life clinical education. Acad Med 2015 Mar;90(3):314-316. [doi: 10.1097/ACM.0000000000000620] [Medline: 25551855]
53. Albrecht U, von JU, Kuebler J, Zoeller C, Lacher M, Muensterer OJ, et al. Google Glass for documentation of medical findings: evaluation in forensic medicine. J Med Internet Res 2014 Feb 12;16(2):e53 [FREE Full text] [doi: 10.2196/jmir.3225] [Medline: 24521935]
54. Aldaz G, Shluzas LA, Pickham D, Eris O, Sadler J, Joshi S, et al. Hands-free image capture, data tagging and transfer using Google Glass: a pilot study for improved wound care management. PLoS One 2015;10(4):e0121179 [FREE Full text] [doi: 10.1371/journal.pone.0121179] [Medline: 25902061]
55. Chai PR, Babu KM, Boyer EW. The Feasibility and Acceptability of Google Glass for Teletoxicology Consults. J Med Toxicol 2015 Sep;11(3):283-287 [FREE Full text] [doi: 10.1007/s13181-015-0495-7] [Medline: 26245879]
56. Chai PR, Wu RY, Ranney ML, Bird J, Chai S, Zink B, et al. Feasibility and Acceptability of Google Glass for Emergency Department Dermatology Consultations. JAMA Dermatol 2015 Jul;151(7):794-796. [doi: 10.1001/jamadermatol.2015.0248] [Medline: 25874898]
57. Muensterer OJ, Lacher M, Zoeller C, Bronstein M, Kübler J. Google Glass in pediatric surgery: an exploratory study. Int J Surg 2014;12(4):281-289 [FREE Full text] [doi: 10.1016/j.ijsu.2014.02.003] [Medline: 24534776]
58. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
59. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care (Internet). York, UK; 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [WebCite Cache ID 6pJ0smh62]
60. Higgins JPT, Green S. The Cochrane Collaboration. 2011. URL: http://www.cochrane-handbook.org [WebCite Cache ID 6pIyMSTKD]
61. Bartolucci AW. Overview, strengths, limitations of systematic reviews and meta-analyses. In: Chiappelli F, editor. Evidence-Based Practice: Toward Optimizing Clinical Outcomes. Berlin, Germany: Springer-Verlag; 2010:34.
62. Dubben H, Beck-Bornholdt H. Systematic review of publication bias in studies on publication bias. BMJ 2005 Aug 20;331(7514):433-434 [FREE Full text] [doi: 10.1136/bmj.38478.497164.F7] [Medline: 15937056]

Abbreviations

ASD: autism spectrum disorder
ECG: electrocardiogram
GG: Google Glass
PD: Parkinson's Disease
RDT: rapid diagnostic test
SAD: social anxiety disorder
SP: standardized patient

Edited by G Eysenbach; submitted 08.08.17; peer-reviewed by N Shah, UV Albrecht; comments to author 24.08.17; revised version received 28.08.17; accepted 28.08.17; published 19.10.17.

Please cite as:
Dougherty B, Badawy SM
Using Google Glass in Nonsurgical Medical Settings: Systematic Review
JMIR Mhealth Uhealth 2017;5(10):e159
URL: http://mhealth.jmir.org/2017/10/e159/
doi:10.2196/mhealth.8671
PMID:29051136

Original Paper

Mobile Phone-Based Measures of Activity, Step Count, and Gait Speed: Results From a Study of Older Ambulatory Adults in a Naturalistic Setting

Cassia Rye Hanton[1*], BS; Yong-Jun Kwon[2*], BS; Thawda Aung[2], BS; Jackie Whittington[1]; Robin R High[3], MBA, MA; Evan H Goulding[4], MD, PhD; A Katrin Schenk[2], PhD; Stephen J Bonasera[1], MD, PhD

[1]Department of Internal Medicine, Division of Geriatrics, University of Nebraska Medical Center, Omaha, NE, United States

[2]Department of Physics, Randolph College, Lynchburg, VA, United States

[3]Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, NE, United States

[4]Department of Psychiatry and Behavioral Sciences, Northwestern University, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Stephen J Bonasera, MD, PhD
Department of Internal Medicine, Division of Geriatrics
University of Nebraska Medical Center
986155 Nebraska Medical Center
Omaha, NE, 68198-6155
United States
Phone: 1 402 559 8409
Fax: 1 402 559 7406
Email: sbonasera@unmc.edu

Abstract

Background: Cellular mobile telephone technology shows much promise for delivering and evaluating healthcare interventions in cost-effective manners with minimal barriers to access. There is little data demonstrating that these devices can accurately measure clinically important aspects of individual functional status in naturalistic environments outside of the laboratory.

Objective: The objective of this study was to demonstrate that data derived from ubiquitous mobile phone technology, using algorithms developed and previously validated by our lab in a controlled setting, can be employed to continuously and noninvasively measure aspects of participant (subject) health status including step counts, gait speed, and activity level, in a naturalistic community setting. A second objective was to compare our mobile phone-based data against current standard survey-based gait instruments and clinical physical performance measures in order to determine whether they measured similar or independent constructs.

Methods: A total of 43 ambulatory, independently dwelling older adults were recruited from Nebraska Medicine, including 25 (58%, 25/43) healthy control individuals from our Engage Wellness Center and 18 (42%, 18/43) functionally impaired, cognitively intact individuals (who met at least 3 of 5 criteria for frailty) from our ambulatory Geriatrics Clinic. The following previously-validated surveys were obtained on study day 1: (1) Late Life Function and Disability Instrument (LLFDI); (2) Survey of Activities and Fear of Falling in the Elderly (SAFFE); (3) Patient Reported Outcomes Measurement Information System (PROMIS), short form version 1.0 Physical Function 10a (PROMIS-PF); and (4) PROMIS Global Health, short form version 1.1 (PROMIS-GH). In addition, clinical physical performance measurements of frailty (10 foot Get up and Go, 4 Meter walk, and Figure-of-8 Walk [F8W]) were also obtained. These metrics were compared to our mobile phone-based metrics collected from the participants in the community over a 24-hour period occurring within 1 week of the initial assessment.

Results: We identified statistically significant differences between functionally intact and frail participants in mobile phone-derived measures of percent activity (P=.002, t test), active versus inactive status (P=.02, t test), average step counts (P<.001, repeated measures analysis of variance [ANOVA]) and gait speed (P<.001, t test). In functionally intact individuals, the above mobile phone metrics assessed aspects of functional status independent (Bland-Altman and correlation analysis) of both survey- and/or performance battery-based functional measures. In contrast, in frail individuals, the above mobile phone metrics correlated with submeasures of both SAFFE and PROMIS-GH.

Conclusions: Continuous mobile phone-based measures of participant community activity and mobility strongly differentiate between persons with intact functional status and persons with a frailty phenotype. These measures assess dimensions of functional status independent of those measured using current validated questionnaires and physical performance assessments to identify functional compromise. Mobile phone-based gait measures may provide a more readily accessible and less-time consuming measure of gait, while further providing clinicians with longitudinal gait measures that are currently difficult to obtain.

(JMIR Mhealth Uhealth 2017;5(10):e104) doi:10.2196/mhealth.5090

KEYWORDS

mobile phone; functional status; mobility; gait speed; mobility measures; LLFDI; SAFFE; PROMIS short; PROMIS Global; step count; behavioral classification; frailty phenotype; normal aging

Introduction

Across a variety of medical disciplines, longer-term measures of gait performance have the potential to benefit both patients and practitioners. Gait speed remains an underutilized clinical measure, despite convincing data suggesting that decreases in gait speed are associated with greater mortality [1], diminished cognition [2], greater functional disability [3], poorer quality of life, and increased healthcare spending [3,4]. There is also evidence suggesting that improved gait speed may be a sensitive biomarker for improved overall functional status [5]. Resistance to including gait speed in current clinical practice is multifactorial, with time and space constraints and provider unfamiliarity being major factors [6]. Longitudinal clinical measures of gait speed are also challenging to obtain, since collecting these measures may be more subject to various biases than more easily obtained metrics such as pulse oximetry or body weight [7].

In the past, gait speed studies have typically relied on measurements taken in the clinic. The standard method for determining gait speed involves timing an individual while walking a short, predetermined distance (eg, 4 to 6 meters). This approach is less than ideal because physical activity, including gait, is influenced by performance biases (eg, participants who know they are being observed try to improve their usual performance), as well as ultradian, circadian, and seasonal changes that cannot be evaluated during a single clinic visit [8]. Furthermore, in older persons, gait speed declines slowly over long periods of time, necessitating repeat observations [9,10].

The rise of widely used electronic devices, such as mobile phones with app capabilities (smartphones), offers great potential for remote monitoring of patient gait speed and other clinically relevant health parameters. We have shown the feasibility of using mobile phone technology to measure an individual's activity and lifespace (eg, the geographic expanse of an individual's day-to-day travels) over prolonged periods of time in a non-invasive, near-continuous, robust, inexpensive, and user-friendly manner [11]. In order to extrapolate health parameters [12] from our participant-derived mobile phone data, we designed algorithms to measure clinically relevant aspects of activity, including gait bout duration, gait speed, and step counts [13]. Additional studies showed that for a broad group of individuals (ranging in age from 21 to 84 years), the activity metrics we measured by this approach strongly correlated with gait speed under controlled laboratory conditions [14].

Here, we show for the first time that mobile phones can provide both continuous and aggregate measures of clinically relevant gait and mobility parameters, including gait speed, step count, and overall activity status, in a community dwelling population going about their day-to-day lives. We gave participants a mobile phone, instructed them in its use, and recorded their activities over the next 24 hours. Validated algorithms were used to classify this data into clinically relevant gait parameters. We studied both healthy (eg, functionally intact) and frail (eg, functionally impaired) community dwelling older individuals. Frailty is a clinical syndrome characterized by poor activity tolerance, weakness, and weight loss unexplained by known diseases of the muscle and brain; frailty is a particularly significant problem leading to greater morbidity and mortality and poorer quality of life [15,16]. Our results suggested that mobile phone-derived measures of these parameters differentiated between older adults without functional limitations and older adults with a frailty phenotype. These mobile phone-derived measures also assessed aspects of functional status distinct from those quantified by either a number of validated questionnaire tools or standard clinical physical performance measures.

Methods

Participant Enrollment Procedures

Participants for this study were recruited from the University of Nebraska Medical Center (UNMC) Geriatrics Clinic and the Engage Wellness Center, both part of UNMC's Home Instead Center for Successful Aging (HICSA). We assembled 2 ambulatory cohorts: one of healthy older individuals with no functional impairment (n=25), and one of frail [17] older individuals (n=18). For our functionally intact group, inclusion criteria were (1) age 55 or older; (2) community dwelling; (3) no serious uncontrolled medical or psychiatric comorbidities; and (4) a minimum score of 23 out of 30 on the Mini-Mental State Examination (MMSE) [18] or Montreal Cognitive Assessment (MoCA) [19]. For our frail group, inclusion criteria also required having 3 of the 5 following clinical conditions present at enrollment: (1) less than 10% unintentional weight loss or body mass index (BMI) less than 18.5 kg/m^2; (2) slow (less than 0.8 m/s) walking speed [20]; (3) weak grip strength (measured by a hand dynamometer, JAMAR, Bolingbrook, IL); (4) reports of exhaustion; and (5) low activity. Of note, the cognitive criteria required that we screened a large number of potential participants for our frailty group. The UNMC Institutional Review Board approved this study. Written

informed consent was obtained from all participants. The enrollment flow diagram is shown in Figure 1 and the baseline cohort characteristics are shown in Table 1.

Self-Reported Functional Status

We used previously validated survey instruments to determine participant self-perceived functional status. These instruments included the (1) functional component of the Late Life Function and Disability Instrument (LLFDI), a comprehensive assessment of function and disability for use in community-dwelling older adults that evaluates self-reported difficulty performing 32 physical activities (eg, use of a stepstool or running to catch a bus), where higher scores indicate higher functional status) [21]; (2) the Survey of Activities and Fear of Falling in the Elderly (SAFFE), a questionnaire evaluating fears associated with performing 11 activities of everyday life (eg, if the participant is limited going to the store or going out when it is slippery) necessary for independent living [22]; (3) Patient Reported Outcomes Measurement Information System (PROMIS) Global Health (GH), short form version 1.1 [23], subdivided into assessments of physical health (PROMIS-PH) and mental health (PROMIS-MH); and (4) PROMIS short form version 1.0 Physical Function 10a (PROMIS-PF). These PROMIS outcome measures were designed to assess patient experience of health outcomes such as pain, fatigue, physical function, depression, anxiety, and social function [24,25]. PROMIS instruments are based on strong psychometrics and consequently have fewer problems with floor and ceiling effect than other survey instruments.

Participants were comfortably seated in a quiet room to complete the above questionnaires, which were administered by a tablet computer (iPad, Apple Inc.). All questionnaire results were stored using the Research Electronic Data Capture (REDCap) database [26]. Participants were given as much time as they needed to complete the surveys. Staff provided no assistance during this process and participants had to complete all the questions to remain eligible for the study.

Figure 1. Enrollment flow diagram.

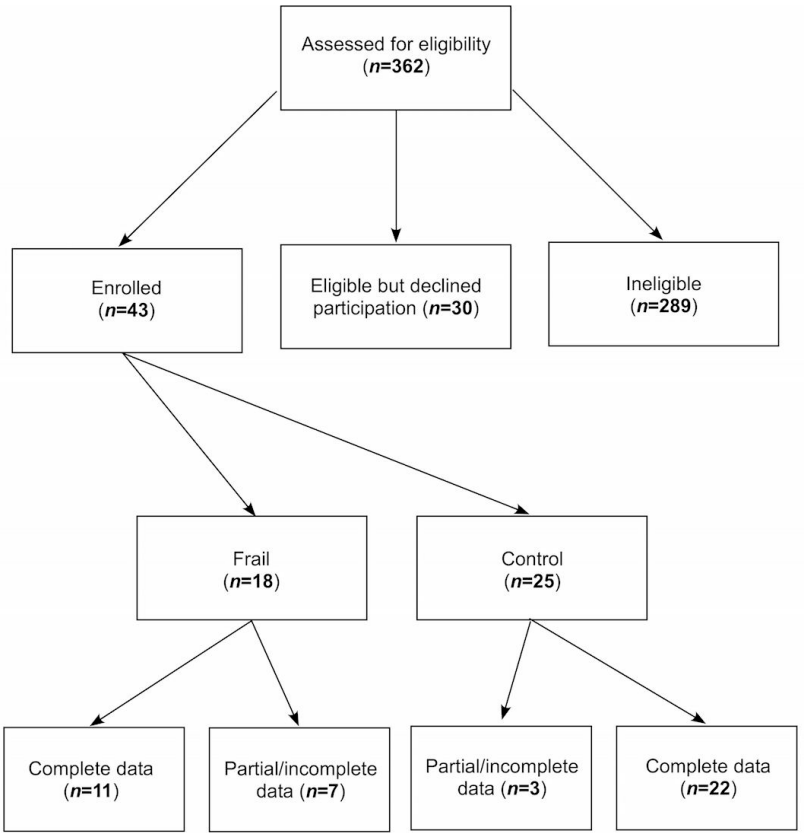


Table 1. Baseline cohort characteristics (N=40).

Characteristic		Functionally-intact, n (%)	Frail, n (%)	*P*
Overall		22 (61%)	18 (39%)	
Gender				.606
	Female	17 (77%)	12 (67%)	
	Male	5 (23%)	6 (33%)	
Age, years				<.001
	50-60	1 (5%)	N/A	
	61-70	5 (23%)	1 (6%)	
	71-80	14 (63%)	5 (28%)	
	81-90	2 (9%)	7 (39%)	
	91-100	N/A	5 (28%)	
Ethnicity				.37
	Non-Hispanic white	20 (91%)	17 (94%)	
	Other	2 (9%)	1 (6%)	
Current residence				
	Home (rented or owned)	19 (86%)	13 (72%)	.26
	Apartment	2 (9%)	4 (22%)	.25
	Assisted living facility	N/A	1 (6%)	.33
	Other	1 (5%)	N/A	.33
Living with				
	Alone	14 (64%)	8 (44%)	.17
	Spouse or significant other	6 (27%)	4 (22%)	.80
	Spouse with children, caregiver	N/A	4 (22%)	.04
	Other	2 (9%)	2 (11%)	.80
Education				<.001
	Grades 9-11	1 (5%)	2 (11%)	
	Grade 12 or GED[a]	2 (9%)	5 (28%)	
	College 1-3 years	8 (36%)	5 (28%)	
	College 4 years	5 (23%)	4 (22%)	
	Graduate school	6 (27%)	2 (11%)	
BMI[b]				.934
	<20	1 (5%)	1 (6%)	
	20-25	6 (27%)	4 (22%)	
	26-30	9 (41%)	5 (28%)	
	31-35	5 (23%)	6 (33%)	
	35+	1 (5%)	2 (11%)	

[a]GED: general education development.

[b]BMI: body mass index.

Clinical Gait Measures

All participants underwent a 4 meter walking test [27] consisting of a 1 meter untimed startup followed by a 4 meter timed evaluation. Participants were given the instruction to "walk at your usual speed" and were permitted to use an assistive device such as a walker or cane at their discretion. Participants then performed a 10 foot "Get Up and Go" test [28]. They began the test seated with their back against the backrest of an armless chair. They were instructed to stand up and "walk at your usual speed" to a mark 10 feet directly in front of the chair, turn around, return to the chair, and sit down again. Timing stopped when their back once again touched the backrest of the chair. Finally, we video recorded participant performance during a Figure-of-8 Walk (F8W) [29]. The camera focused on the participant's lower legs and feet during the test. No identifying features were photographed. Participants were instructed to walk in a figure-of-8 at their self-selected pace around 2 cones placed 5 feet apart. Total completion times, the number of steps to complete the F8W, and gait smoothness were recorded. Two trials of all physical assessment tests were performed. Participants tolerated all of these clinical assays with ease.

Gait Data Acquisition

Nokia N79 mobile phones (White Plains, NY) with an intrinsic three-dimensional (3D) accelerometer were used to measure mobility and locomotion for extended periods of time in community dwelling individuals of both cohorts. Acceleration values were sampled and written to memory using custom Python software (Python for S60 v1.9.7) [30], running on a Symbian S60 V3FP2 OS (San Francisco, CA). The mobile phone was placed in either the participant's right or left pocket, over the hip, and the location recorded. A previous study of ours showed that location did not impact data collection [14].

Protocol

Participants were fitted with a mobile phone; the proper use and correct placement of this device was demonstrated. Participants were instructed to wear the mobile phone for the next 24 hours, except when bathing, swimming, or sleeping.

Participants were then briefly videotaped walking on a treadmill (SCIFIT, Tulsa, OK) for 5 minutes at 2 mi/hr or a speed more comfortable for the individual participant. Participants unable to walk on the treadmill due to limited mobility, the need for assistive devices, or other factors, were not asked to complete this portion of the study. Gait speed calculations depended upon stride length, which we derived from treadmill locomotion videos (1.38 m functionally intact; 0.83 m frail).

Data Quality Control and Classification

Survey data was scored per instrument instructions. For PROMIS measures, *t* scores were determined from raw scores by appropriate conversion tables. Raw acceleration data was low-pass filtered and baseline acceleration normalized to 1 g over the entire duration of data collection [13]. Our classification algorithm first identified epochs of "forgotten phone" versus epochs of participant carrying the phone. For epochs of participant carrying the phone, we then classified behavior into active or inactive states, using a windowed (68 s long) Fourier analysis approach [13,31]. Active states were further differentiated into states with minimal locomotion, states with ongoing locomotion, and states where the participants were climbing stairs. Ongoing locomotion was then quantified for step count and gait speed. Gait speed calculations depended on treadmill video-derived values of stride length.

Statistical Analysis

Step count, gait speed, and activity count were primary outcomes with cohort (functionally intact versus impaired) and time as factors. All comparisons not involving time were performed by 2-tailed *t* tests assuming unequal variances. Multiple comparisons were adjusted using the Bonferroni technique. For comparisons over time, we performed repeated measures analysis of variance (ANOVA) including all interaction terms; interactions not found to be significant were dropped from later models. All post hoc testing was performed using Tukey test. To measure pairwise agreements between mobile phones, surveys, and physical performance battery-based functional measures, we evaluated Bland-Altman plots using MATLAB (blandaltman.m). For gait speed measures, we performed a bootstrap analysis using MATLAB (datasample) to determine statistical significance between the functionally intact and frail cohorts. Functional questionnaire data and clinical physical performance measures were analyzed by 1-way ANOVA. Spearman correlations were determined to assess agreements between mobile phone-based measures and the survey- and/or performance-based metrics. Finally, cohort demographic factors were compared using independent samples *t* test assuming equal variances (2-tailed). All analyses were performed using SPSS (IBM SPSS Statistics 22.0, Armonk, New York, USA) or MATLAB (R2011b, MathWorks, Natick, MA).

Results

A total of 362 participants were assessed for study eligibility (Figure 1) and, from those, 73 (20.2%, 73/362) were identified as potential study participants. Of those, 30 (41%, 30/73) individuals declined participation and 43 (59%, 43/73) were consented into the study. Of the 43 participants enrolled in the study, 25 (58%, 25/43) were enrolled in the control (functionally intact) arm and 18 (42%, 18/43) were enrolled in the frail (functionally impaired) arm. Unfortunately, mobile phone sensor data was not collected from 10 participants (23%, 10/43) in the functionally intact arm and 7 (16%, 7/43) from the frail arm due to technical problems with data transmission and storage. By far, the greatest challenge we encountered during participant recruitment was identifying frail individuals with preserved cognition.

During our 24-hour study period, all 22 functionally-intact participants recorded at least 14 hours of data, with a mean of 17.3 hours (range 14 to 20 hours) for a total of 380 hours suitable for analysis. We determined that 11 (61%, 11/18) participants from the frail arm recorded at least 9 hours of data, with a mean of 19.9 hours (range 9 to 24 hours) for a total of 210 hours, of which 209 were suitable for analysis (1 hour prematurely truncated). There was no significant difference in the number of hours recorded for the functionally intact versus frail cohort (P=.17) when normalized over the 24-hour day. We obtained similar amounts of data from both functionally intact

and frail older individuals. Baseline demographics between the 2 groups revealed that our functionally intact participants were younger and had higher educational achievement compared to our frail participants; cohorts were otherwise comparable regarding gender, ethnicity, housing, and BMI (Table 1).

Gait Assessment Survey Instruments and Clinical Performance Measures

We chose our questionnaires and performance assessments based on prior validation, current clinical and/or research use, and face validity. In total, 40 (93%, 40/43) participants successfully completed the questionnaires and physical performance batteries. As expected, our analysis demonstrated that LLFDI, SAFFE, PROMIS-PH, and PROMIS-PF all differentiated individuals from functionally intact and frail groups (Table 2). Our 3 clinical physical performance measures (10-foot timed Get up and Go, 4 Meter Walk, and F8W) also performed as expected (Table 2), with all measures demonstrating robust differences between our functionally intact and frail cohorts. The LLFDI scores range between 0 (full limitations for performing tasks) to 100 (no limitations for performing tasks) [32]. SAFFE activity level scores ranged between 0 (lowest function) and 11 (highest function), SAFFE fear of falling scores range between 0 (no fear of falling) and 3 (high fear of falling), and SAFFE activity restriction scores range between 0 (no activity restrictions) and 11 (marked activity restrictions) [33]. PROMIS *t* scores set average performance for a US-based population at 50 (SD 10 points), with better function indicated by higher scores. Scoring was performed per PROMIS-GH and PROMIS-PF [34]. EuroQol scores were derived from PROMIS-GH, and range between 0 (very poor health-related quality of life) and 1 (very high health-related quality of life).

Table 2. Statistical significance of standard questionnaire and physical performance battery in discriminating functionally impaired from functionally intact participants.

Survey instrument	Intact[a]	Impaired[a]	*P* [b]
Questionnaire			
LLFDI[c] Overall Function	64.04	46.91	<.001
LLFDI basic lower extremity function	76.02	54.23	<.001
LLFDI advance lower extremity function	56.15	22.41	<.001
LLFDI upper extremity function	79.57	68.95	.257
SAFFE[d] activity level	9.32	6.38	<.001
SAFFE fear of falling	0.24	0.39	.052
SAFFE activity restriction	2.45	6.59	<.001
PROMIS[e]-PF[f]	48.94	34.43	<.001
PROMIS-PH[g]	51.99	40.52	<.001
PROMIS-MH[h]	63.65	55.79	.042
EuroQol	0.764	0.64	<.001
Performance battery[i]			
Timed Get Up and Go (10 ft)	10.63	21.79	.003
4 Meter Walk	4.30	10.66	.004
Figure-of-8 Walk	9.19	19.28	.008

[a]Mean performance.

[b]*P* values are 2-sided *t* test, unequal variance, with Bonferroni correction.

[c]LLFDI: Late Life Function and Disability Instrument.

[d]SAFFE: Survey of Activities and Fear of Falling in the Elderly.

[e]PROMIS: Patient Reported Outcomes Measurement Information System.

[f]PROMIS-PF: PROMIS Physical Function.

[g]PROMIS-PH: PROMIS Global Physical Health.

[h]PROMIS-MH: PROMIS Global Mental Health.

[i]Values for all performance battery measures are reported in seconds.

Figure 2. Mobile phone-derived activity metrics discriminate between frail and functionally-intact individuals. 24-hour time budget for functionallyintact (left) and functionally impaired (right) participants. Time spent in active state (blue slices) is further broken down into periods of low (brown)and high (red, green) physical activity. Percentages (bold) statistically differ between cohorts.

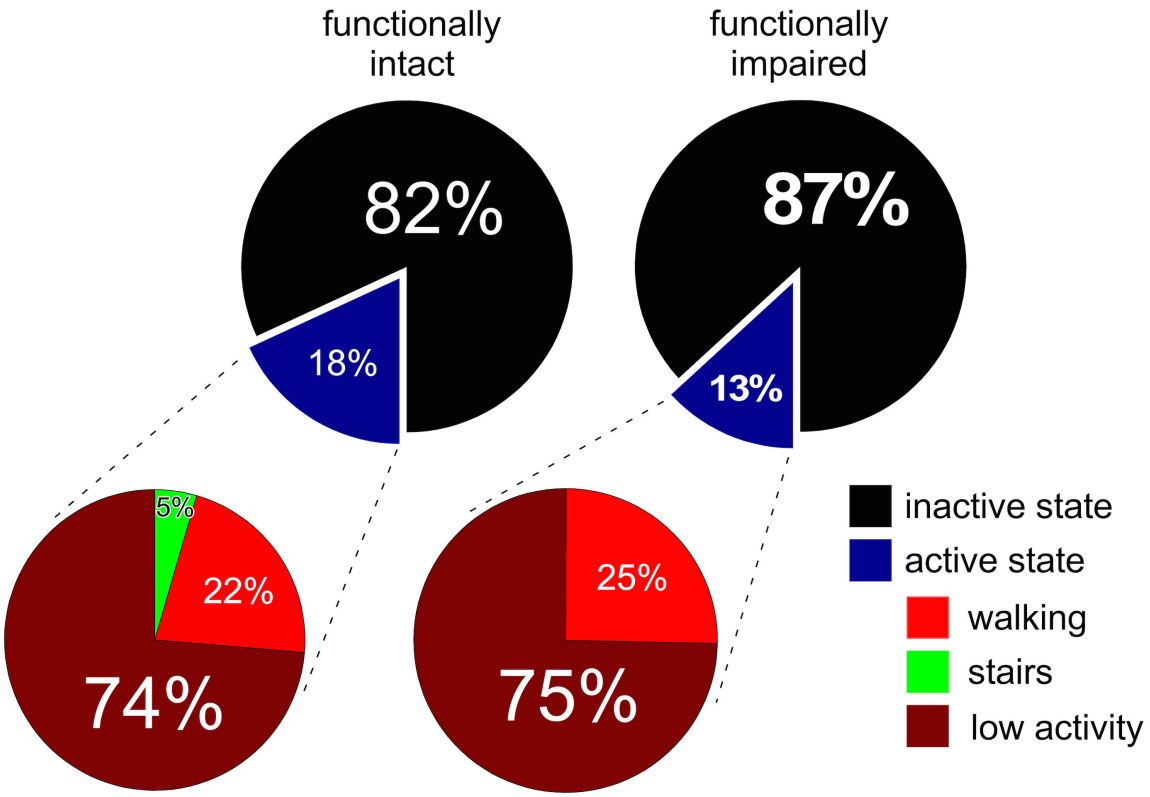
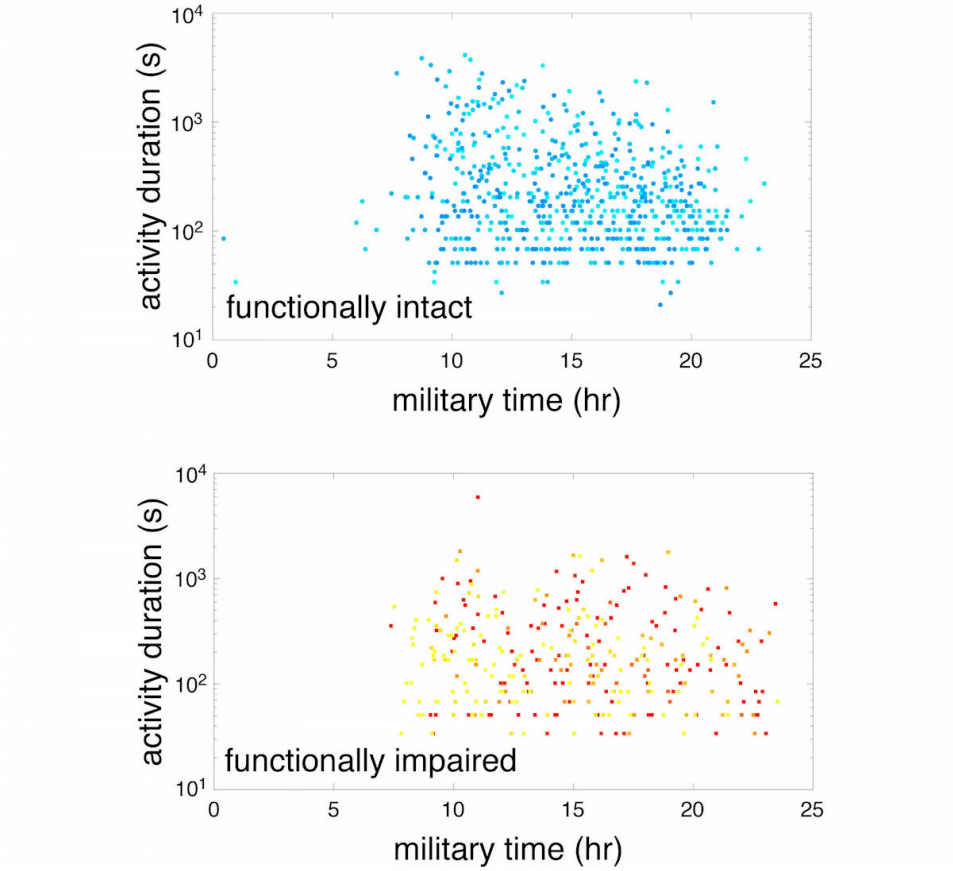


Figure 3. Semilog (y axis) of activity duration versus start time of that particular activity bout (x axis) in functionally intact individuals (upper) and functionally impaired individuals (lower). Each point represents a specific activity bout and each color corresponds to a specific participant ("cool" or "warm" colormap).



Mobile Phone-Based Functional Measures

After determining that our questionnaire-based measures and physical performance battery successfully differentiated functionally intact from functionally impaired individuals, we examined whether our mobile-phone-based measures of physical activity did so as well. Active states were defined as periods where the participant was walking, climbing stairs, or otherwise active (high physical activity classification) [13]. Inactive states were defined as when the participant was resting or driving (low physical activity) [13]. We noted significant differences in participant 24-hour and active state time budgets (Figure 2). Overall, the functionally intact group were active approximately 18% of the day with a mean of 18.13% (SD 5.54%); while the frail group displayed significantly less activity with a mean of 13.19% (SD 5.20%; *P*=.02 for intact versus frail groups, 2-sided *t* test). There were no phenotypic differences in active state onset rate between functionally intact and impaired individuals with 2.63 onsets/hr (SD 0.162) and 2.48 onsets/hr (SD 0.219), respectively (*P*=.60, 2-sided Student *t* test). Functionally intact individuals had longer active state durations of 373.85 s (SD 20.66) compared to frail individuals with active state durations of 300.19 s (SD 25.79; *P*=.04; 2-sided Student *t* test) (Figure 3). Similarly, average gait speed (measured over a 24-hour window) differed significantly between frail and functionally intact groups with mean gait speeds of 0.76 m/s (SD 0.08) and 1.22 m/s (SD 0.14), respectively ($F_{1,30}$=21.1, *P*<.001) (Figure 4).

The average number of step counts throughout a circadian day also differed between frail and functionally intact groups (Figure 5). Repeated measures 2-way ANOVA, with log step count as the dependent variable and functional status and time as independent variables, found functional status, time, and the status by time interaction to be significant ($F_{1,30}$=12.1, *P*=.002 for functional status; $F_{23,521}$=9.0, *P*<.001 for time; $F_{23,521}$=1.6, *P*=.045 for functional status by time interaction). Overall, all mobile phone collected outcomes, including step count, gait speed, activity classification, and percent activity were statistically significant in our study, indicating substantial differences between functionally intact and frail participants.

Aspects of Gait Assessed by Survey and/or Performance Battery and Mobile Phone Functional Measures Assess Different Aspects of Gait

We further decided to determine (1) if our mobile phone-based measures identified similar elements of frailty as survey and physical performance assessments; and (2) if performance data obtained in this study (whether from mobile phone, survey, or performance battery) demonstrated internal consistency. First, we evaluated mean-difference (Bland-Altman) plots in a pairwise manner comparing mobile phone-, survey-, and

performance battery-based functional measures. Bland-Altman plots provide a graphical approach to determine if results from two different measurement methods assessed a similar construct; if this were the case, the plotted residuals would form a relatively uniform-width band parallel to the x axis. Our analysis suggested that only LLFDI overall/LLFDI basic and PROMIS PF/PROMIS PH measured similar outcomes in both functionally intact and frail participants (Multimedia Appendix 1).

Figure 4. Mean daily gait speed histogram depicting significant differences between functionally intact (blue) and functionally impaired (red) participants. Bootstrap estimates of mean gait speed are provided behind data histograms (light red for functionally impaired; estimate for functionally intact group is completely behind data histogram).

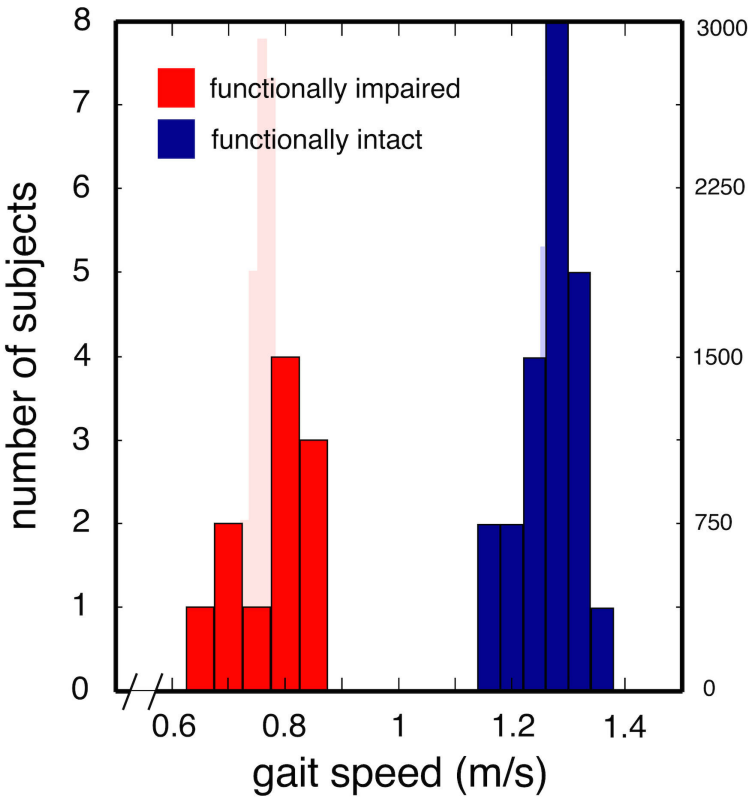


Figure 5. Step count versus circadian time for functionally intact (blue) and frail (red) individuals. Bars are plus or minus one standard error of the mean. Time values given in military time.

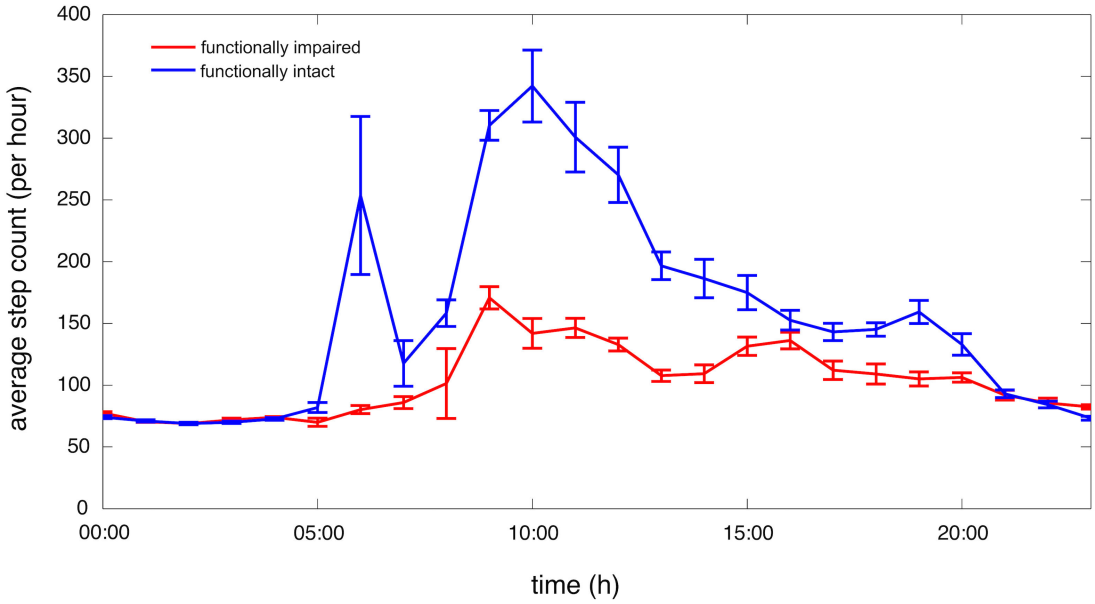
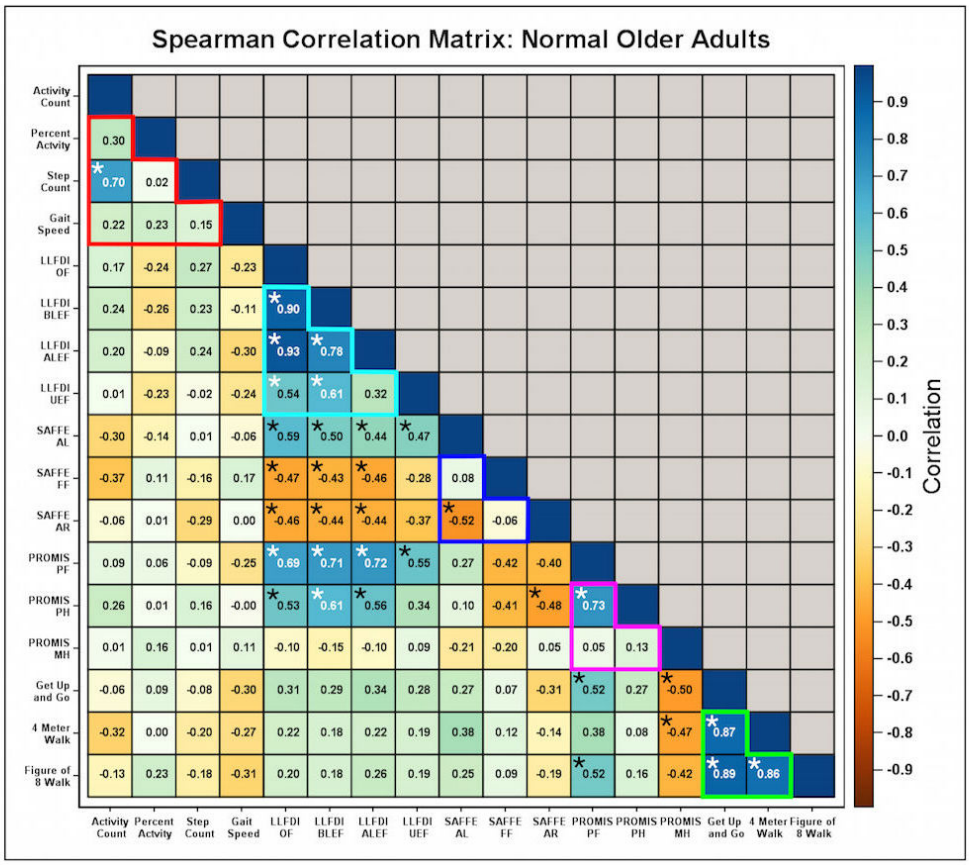
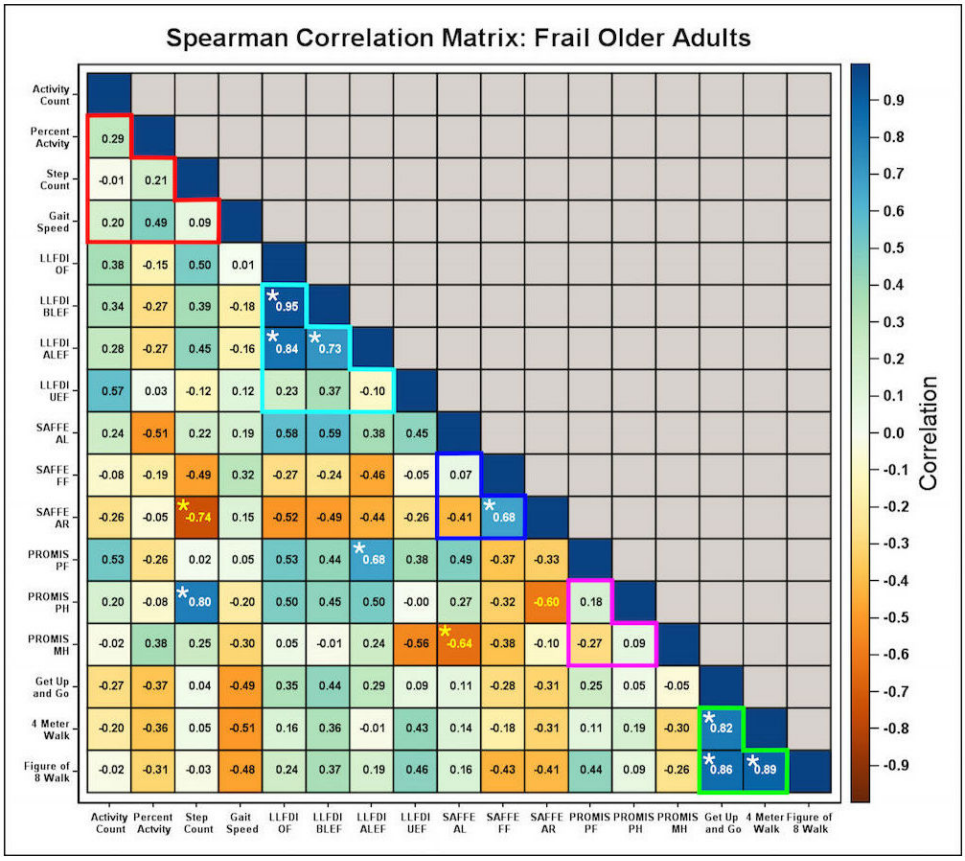


Figure 6. Multiple significant correlations across different functional assessment metrics are present in functionally intact older adults. Confusion matrix depicting correlation structure (metrics across matrix bottom row) of mobile phone-based activity measures (activity count, percent activity, step count, gait speed), questionnaire-based measures of functional status (LLFDI, SAFFE, PROMIS), and performance battery based measures of functional status (Get Up and Go, 4 meter walk, F8W) for functionally intact participants. For each entry, correlation strength is depicted as the color within the box; interpretation color bar provided on the right. Numeric values within each box are individual correlations. Asterisks depict interactions with P values less than .01. Interactions grouped within the red lines depict correlations within mobile phone-derived activity measures; interactions grouped within cyan lines depict correlations within LLFDI measures; interactions grouped within blue lines depict correlations within SAFFE measures; interactions within grouped violet lines depict correlations within PROMIS measures; interactions grouped within green lines depict correlations within performance battery measures.



We then calculated Spearman correlations between survey and/or performance instruments and the mobile phone-based functional measurements for functionally intact (Figure 6) and frail (Figure 7) participants. In functionally intact participants, we noted significant within-test correlations for our mobile phone-based monitoring metrics (step and activity count; 1 of 6 potential correlations), all LLFDI metrics (except for those measuring upper extremity function [UEF]; 5 of 6 potential correlations), SAFFE metrics (activity restriction and limitation; 1 of 3 potential correlations), PROMIS metrics (PROMIS-PH and PROMIS-PF; 1 of 3 potential correlations), and all performance battery results (3 of 3 potential correlations). LLFDI metrics (except UEF) also strongly correlated with results from both SAFFE and PROMIS (except PROMIS-MH; 17 of 25 potential correlations). By contrast, both within-instrument and across-instrument correlations were weaker in frail adults with functional impairment. Only performance battery and subsets of LLFDI scores remained significantly correlated with one another (3 of 3 potential correlations for functional battery metrics; 3 of 6 potential correlations for LLFDI metrics). Much of the correlation between LLFDI and SAFFE/PROMIS metrics was no longer observed. Step and activity counts no longer correlated with one another in functionally impaired individuals; however, step count now demonstrated significant correlations with both SAFFE activity restriction and PROMIS-PH.

Figure 7. Fewer significant correlations across different functional assessment metrics are present in functionally impaired older adults. Confusion matrix depicting correlation structure of mobile phone-based activity measures, questionnaire-based measures of functional status, and performance battery based measures of functional status for frail participants. Layout similar to Figure 6.



Discussion

Principal Findings

To our knowledge, we present the first demonstration that mobile phones measure clinically relevant functional metrics, including overall activity, gait speed, and step count. These measures were taken over one day in naturalistic conditions and real-life settings, and thus provided insights regarding individual function outside of the clinic. We provided further validation of the LLFDI, SAFFE, PROMIS-GH, PROMIS-PF, timed 4-meter walking test, timed Get Up and Go, and F8W assays demonstrating that cognitively intact individuals with functional loss had worse performance on all of these assays compared to functionally intact individuals. In functionally intact individuals, mobile phone-based metrics and survey and/or performance battery results did not strongly correlate with one another, suggesting that these different tools measure distinct aspects of physical function. However, in cognitively intact individuals with functional loss, mobile phone-based functional metrics strongly correlated with components of both SAFFE and PROMIS. Thus, in functionally impaired individuals, mobile phone-based metrics of impaired physical function reflected parallel losses of both perceived and enabled physical function.

Measuring Individual Functional Status Using Mobile Phones

Our study advances the goal of an easy-to-use, robust, accurate, second nature system that measures clinically relevant activity metrics (onsets, durations, step counts, and gait speeds) in different ambulatory populations. This goal is attainable with appropriate hardware and software. For example, over 50 years ago, Stunkard [35] showed the feasibility of using pedometers to estimate individual walking distance over long observations. Technical refinements (improved accelerometer technology, device durability, device data logging) have since increased data accuracy and temporal precision [36-38]. However, dedicated devices validated in many small trials to measure individual activity status have not "caught on" with the population at large, potentially because these devices did not successfully address human usability factors [39]. By contrast, mobile phones have become a nearly ubiquitous technology [40-42]. This is particularly true among younger and middle-aged adults, whose quality of life stands to significantly benefit from advances in mobile phone-based healthcare delivery and follow-up. However, we do note that while we had excellent adherence to our data collection in the functionally intact older adult group, we had less success with data collection in the frail older adult group. This decrease in adherence suggests that older adults with functional limitations may have more difficulties using this technology successfully. Devices to serve this

population may require further engineering to optimize user interface features.

Metric Validation

This study provided additional opportunity to further validate a number of questionnaire and performance instruments designed to measure functional status. The LLFDI [21] assesses two distinct outcomes: function (ability to do discrete actions or activities), and disability (performance of socially defined life tasks). Prior studies have validated LLFDI for identifying functional deficits in independent older adults [43], institutionalized older adults [44], older adults with knee osteoarthritis [45], older adults with chronic renal disease [46] and incontinence [47], and persons undergoing cardiac physical therapy [48]. LLFDI has comparable psychometric properties to performance-based measures of upper and lower extremity function [49]. Our results suggested that LLFDI can discriminate functional status between a cohort of functionally intact older adults and persons meeting frailty criteria without cognitive impairment. We also demonstrated that in functionally intact, but not frail, individuals, LLDFI is highly correlated across its functional submeasures (LLFDI basic lower extremity function [BLEF], LLFDI advanced lower extremity function [ALEF], etc), and is significantly correlated to both SAFFE and PROMIS (except PROMIS-MH) scores. For all participants, LLFDI was not significantly correlated to either mobile phone gait speed or physical performance battery measures.

SAFFE evaluates how fear of falling influences participant activity participation or restriction. It has been validated in community dwelling older adults [22,50], older adults with mobility limitations [51], and extensively utilized in studies of persons with Parkinson's disease [52-54] as well as individuals receiving post-fall physical therapy who have a fear of falling [22]. Our results further suggested that SAFFE can discriminate functional status between a cohort of functionally intact older adults and persons meeting frailty criteria (albeit, we did not evaluate balance or falls in any of our participants). As mentioned above, in functionally intact (but not frail) individuals, SAFFE showed significant correlations to both LLFDI and PROMIS (except PROMIS-MH) scores. For all participants, correlations of submeasures within SAFFE (eg, SAFFE FF, SAFFE AL) were weaker ($-.67 < r < -.33$; $.33 < r < .67$). SAFFE scores also did not significantly correlate with either mobile phone gait speed or physical performance battery measures. Previous studies also have demonstrated weak correlation between SAFFE scores and accelerometer-based activity measures [55].

PROMIS-GH evaluates individual physical, mental, and social health domains, and is thus a more all-encompassing view of health status [56]. PROMIS-PF is a shorter, 10-question instrument that assesses individual physical health capacity without requiring a lengthy physical function instrument [23]. Compared to LLFDI and SAFFE, which were developed specifically for use in older populations, PROMIS-GH and PROMIS-PF assessments were developed for general adult populations [23,57]. Both of these instruments have previously been validated in a large, cross-sectional sample of independently dwelling US adults [23,58,59], as well as persons with chronic pelvic pain [60], cancer [61,62], or in preparation for surgical procedures [63]. Our results suggest that PROMIS-PF and PROMIS-PH can discriminate functional status capacity between a cohort of intact older adults and persons meeting frailty criteria. As mentioned above, in functionally intact (but not frail) individuals, we noted significant correlations between PROMIS and both LLDFI and SAFFE measures. PROMIS-GH and PROMIS-PF were also significantly correlated for functionally intact individuals, as were multiple physical performance battery measures. These findings suggested that PROMIS measures of physical capacity accurately reflected observed physical function in functionally intact individuals. However, in frail individuals, PROMIS measures had no significant correlations with all other metrics we quantified except for mobile phone-derived step count, LLFDI ALEF, and SAFFE AL.

A variety of physical performance measures have been adapted for clinic use, including the 4 Meter Walk [64], the timed Get Up and Go test [65], and the F8W test [29]. Both the timed Get Up and Go and F8W tests focus on older populations and have been used to assess community dwelling older adults and individuals with Parkinson's disease. The 4 Meter Walk was developed for persons ranging from 7 to 85 years of age, and is a validated functional measure in persons with peripheral arterial disease [66] and cerebrovascular disease [67]. Our results demonstrated that all of these gait-associated performance batteries reliably distinguished between functionally intact older adults and older adults meeting frailty criteria. We also noted high correlations across these physical performance tests in both functionally intact and frail individuals. However, none of these measures correlated well with our mobile phone-derived activity and gait metrics.

Mobile Phone-Derived Gait Metrics Reflect Different Aspects of Physical Function

In functionally intact individuals, there was little correlation between activity and gait metrics measured by mobile phone and participant responses to the LLFDI, SAFFE, or PROMIS instruments, or to physical battery performance. Similarly, Bland-Altman plots revealed that mobile phone-based metrics of physical activity and gait speed measured different aspects of physical capacity compared to LLFDI, SAFFE, PROMIS, or physical performance batteries. In functionally intact individuals, activity count, daily activity time budget, step count, and gait speed may undergo significant variation within a single individual as well as across many individuals. In other words, these particular functional metrics have considerable dynamic range. By contrast, survey- and physical performance-based instruments are well known to demonstrate ceiling effects in community dwelling individuals [68,69]. Thus, our functionally intact cohort may demonstrate few and weak correlations between mobile phone-based measures of physical activity and survey- or performance battery-based measures of the same, while simultaneously observing more and greater correlations when comparing measures known to have ceiling effects. We observed precisely this finding in our study.

However, functional measures characterizing frail individuals are far less likely to be influenced by ceiling effects. The

decreased dynamic range and increased variability in functional status characteristic of frailty suggest that fewer correlations between different measures of physical capacity should occur in frail individuals. We noted this finding in our study as well. Finally, we noted significant correlations between step count (measured by mobile phone) and SAFFE activity restriction and PROMIS-PF. Mobile phone-derived gait metrics may estimate both activity restrictions and overall physical health (as well as gait speed, step count, and activity status) in older adults as they progress through stages of functional loss and ultimately become functionally impaired.

Limitations

We recognize several limitations in this study, mostly regarding participant characteristics. Our desire to test cognitively intact individuals with functional impairments significantly limited our participant pool. While we ultimately envision that this technology will be used by cognitively impaired persons, for validation purposes we wanted to ensure that group differences could be attributed mostly to functional differences. We did not enroll a large group of cognitively intact individuals with functional deficits; however, given our effect size, we had adequate statistical power for discrimination. Our functionally intact group, self-selected from persons enrolled in a UNMC fitness program, sampled more health literate, financially secure, and higher educated individuals compared to community averages. We also did not quantify additional confounds, including medical comorbidities and pharmacotherapy. However, adjustment of study outcomes for these factors would likely have had only minimal impact on study outcome. Not surprisingly, we continued to note variable participant adherence for keeping the mobile phone during the study. While some participants successfully carried the phone and collected data for almost an entire 24-hour time frame, other individuals wore the phone for 10 hours or less. However, in practice, if individuals were to only collect data for brief, random periods each day, over longer time periods they would produce significant and robust datasets suitable for functional inference.

Given the worldwide ubiquity of mobile phone technology, and decreasing costs associated with mobile phone ownership, this study suggests that future healthcare systems should consider leveraging patient mobile phones to collect data associated with individual functional status (respecting patient privacy and autonomy), develop patient functional exemplars, and refine algorithms that not only calculate activity and gait functional metrics as above, but further identify within-individual acute and subacute functional changes in a reliable, robust, and efficient manner. This approach to population-wide healthcare is in its infancy, but already there is highly promising data suggesting that accurate knowledge of individual day-to-day patterns of behavior and functional status can be used to make rapid and accurate diagnoses of acute disease states [70]. Mobile phones also measure lifespace (an independent metric strongly associated with clinically important healthcare outcomes) [71,72] with high accuracy [73]. Ultimately, integrating these approaches into a comprehensive patient care platform that includes caregiver, decision making, and medication support may lead to significant improvements in patient quality of life, decreased healthcare spending, and improved care outcomes in persons with chronic illnesses, such as Alzheimer's disease.

Acknowledgments

This work was supported by the Alzheimer's Association Everyday Technology for Alzheimer's Care (ETAC) grant 11-206024 (CRH, AKS, SJB), R34MH100460-01 (EHG). The authors would like to thank Samsung USA (R Cha, C Sastry) for their technical and material support, P Moffatt, RN, Jeanne Costello, RN, and Kelli Kubik, RN for their help enrolling participants with functional limitations, and the UNMC EngAge program for their support enrolling functionally intact older adults. Study data were collected and managed using REDCap electronic data capture tools hosted at UNMC. We thank Ran Klein for the blandaltman.m submission to Mathworks Central. We thank William L Lyons, MD, Jane F Potter, MD, and Kenneth E Covinsky, MD for provocative conversations discussing these study results. Most importantly, we thank our study participants for their generous gifts of time, effort, and interest.

Conflicts of Interest

SJB, AKS, and EHG have received a patent regarding aspects of this technological approach (US Patent 9,106,718 B2; Lifespace data collection from discrete areas).

Multimedia Appendix 1

Bland-Altman plots demonstrate that mobile phone-based measures of activity and gait are independent of functional measures obtained from LLFDI, SAFFE, PROMIS PF, PH, and MH, and a physical performance battery. Matrix array depicting pairwise interactions between different study metrics. Matrix divided into regions (depicted by grey lower case letter in background) that group assay agreements for (a) all mobile-based metrics; (b) all LLFDI metrics; (c) all SAFFE metrics; (d) all PROMIS metrics; (e) all physical performance battery metrics; (f) mobile phone and LLFDI metrics; (g) mobile phone and SAFFE metrics; (h) mobile phone and PROMIS metrics; (i) mobile phone and physical performance battery metrics; (j) LLFDI and SAFFE metrics; (k) LLFDI and PROMIS metrics; (l) LLFDI and physical performance battery metrics; (m) SAFFE and PROMIS metrics, (n) SAFFE and physical performance battery metrics; and (o) PROMIS and physical performance battery metrics. Each plot is a mean-difference (Bland-Altman) plot comparing the metric listed at the top of the column to the metric listed at the left of the

row. In each plot, RPC denotes reproducibility coefficient, CV is the coefficient of variation, and dotted lines depict 1.96 standard deviation from mean difference. Frail subjects denoted by red markers; functionally intact subjects denoted by blue markers. Note that with the exception of the LLFDI overall and LLFDI basic comparison, and the PROMIS PF and PROMIS PH comparison, none of the examined metrics appear to assess the same physical function constructs.

[PNG File, 6MB - mhealth_v5i10e104_app1.png]

Multimedia Appendix 2

CONSORT checklist.

[PDF File (Adobe PDF File), 758KB - mhealth_v5i10e104_app2.pdf]

References

1. Pahor M, Guralnik JM, Ambrosius WT, Blair S, Bonds DE, Church TS, LIFE study investigators. Effect of structured physical activity on prevention of major mobility disability in older adults: the LIFE study randomized clinical trial. JAMA 2014 Jun 18;311(23):2387-2396 [FREE Full text] [doi: 10.1001/jama.2014.5616] [Medline: 24866862]
2. Dodge HH, Mattek NC, Austin D, Hayes TL, Kaye JA. In-home walking speeds and variability trajectories associated with mild cognitive impairment. Neurology 2012 Jun 12;78(24):1946-1952 [FREE Full text] [doi: 10.1212/WNL.0b013e318259e1de] [Medline: 22689734]
3. Abellan van Kan G, Rolland Y, Andrieu S, Bauer J, Beauchet O, Bonnefoy M, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people an International Academy on Nutrition and Aging (IANA) Task Force. J Nutr Health Aging 2009 Dec;13(10):881-889. [Medline: 19924348]
4. Hardy SE, Kang Y, Studenski SA, Degenholtz HB. Ability to walk 1/4 mile predicts subsequent disability, mortality, and health care costs. J Gen Intern Med 2011 Feb;26(2):130-135 [FREE Full text] [doi: 10.1007/s11606-010-1543-2] [Medline: 20972641]
5. Wong WL, Masters RS, Maxwell JP, Abernethy AB. Reinvestment and falls in community-dwelling older adults. Neurorehabil Neural Repair 2008;22(4):410-414. [doi: 10.1177/1545968307313510] [Medline: 18334603]
6. Allum JH, Carpenter MG, Horslen BC, Davis JR, Honegger F, Tang K, et al. Improving impaired balance function: real-time versus carry-over effects of prosthetic feedback. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1314-1318. [doi: 10.1109/IEMBS.2011.6090309] [Medline: 22254558]
7. Krebs DE, Edelstein JE, Fishman S. Reliability of observational kinematic gait analysis. Phys Ther 1985 Jul;65(7):1027-1033. [Medline: 3892553]
8. Guyatt GH, Pugsley SO, Sullivan MJ, Thompson PJ, Berman L, Jones NL, et al. Effect of encouragement on walking test performance. Thorax 1984 Nov;39(11):818-822 [FREE Full text] [Medline: 6505988]
9. Atkinson HH, Rosano C, Simonsick EM, Williamson JD, Davis C, Ambrosius WT, Health ABC study. Cognitive function, gait speed decline, and comorbidities: the health, aging and body composition study. J Gerontol A Biol Sci Med Sci 2007 Aug;62(8):844-850. [doi: 10.1093/gerona/62.8.844] [Medline: 17702875]
10. Watson NL, Rosano C, Boudreau RM, Simonsick EM, Ferrucci L, Sutton-Tyrrell K, Health ABC Study. Executive function, memory, and gait speed decline in well-functioning older adults. J Gerontol A Biol Sci Med Sci 2010 Oct;65(10):1093-1100 [FREE Full text] [doi: 10.1093/gerona/glq111] [Medline: 20581339]
11. Schenk AK, Witbrodt BC, Hoarty CA, Carlson RH, Goulding EH, Potter JF, et al. Cellular telephones measure activity and lifespace in community-dwelling adults: proof of principle. J Am Geriatr Soc 2011 Feb;59(2):345-352 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03267.x] [Medline: 21288235]
12. Committee on Leading Health Indicators for Healthy People 2010, Institute of Medicine. Leading Health Indicators for Healthy People 2010: First Interim Report. Washington, DC: National Academies Press; 1998.
13. Kwon Y, Aung T, Synovec S, Oberle A, Rye-Hanton C, Whittington J, et al. BioRxiv. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 2017 Jul 06. Classifying smartphone-based accelerometer data to obtain validated measures of subject activity status, step count, and gait speed URL: http://www.biorxiv.org/content/biorxiv/early/2017/07/06/160317.full.pdf [accessed 2017-07-07] [WebCite Cache ID 6s2DaWPDf]
14. Carlson RH, Huebner DR, Hoarty CA, Whittington J, Haynatzki G, Balas MC, et al. Treadmill gait speeds correlate with physical activity counts measured by cell phone accelerometers. Gait Posture 2012 Jun;36(2):241-248 [FREE Full text] [doi: 10.1016/j.gaitpost.2012.02.025] [Medline: 22475727]
15. Shamliyan T, Talley KM, Ramakrishnan R, Kane RL. Association of frailty with survival: a systematic literature review. Ageing Res Rev 2013 Mar;12(2):719-736. [doi: 10.1016/j.arr.2012.03.001] [Medline: 22426304]
16. Birren JE, Lubben JE, Rowe JC, Deutchman DE. The Concept and Measurement of Quality of Life in the Frail Elderly. San Diego: Academic Press; 1991.
17. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in Older Adults: Evidence for a Phenotype. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2001 Mar 01;56(3):M146-M157. [doi: 10.1093/gerona/56.3.M146] [Medline: 11253156]

18. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. Journal of Psychiatric Research 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
19. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
20. Castell M, Sánchez M, Julián R, Queipo R, Martín S, Otero Á. Frailty prevalence and slow walking speed in persons age 65 and older: implications for primary care. BMC Fam Pract 2013 Jun 19;14:86 [FREE Full text] [doi: 10.1186/1471-2296-14-86] [Medline: 23782891]
21. Sayers SP, Jette AM, Haley SM, Heeren TC, Guralnik JM, Fielding RA. Validation of the Late-Life Function and Disability Instrument. J Am Geriatr Soc 2004 Sep;52(9):1554-1559. [doi: 10.1111/j.1532-5415.2004.52422.x] [Medline: 15341561]
22. Li F, Fisher KJ, Harmer P, McAuley E, Wilson NL. Fear of Falling in Elderly Persons: Association With Falls, Functional Ability, and Quality of Life. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 2003 Sep 01;58(5):P283-P290. [doi: 10.1093/geronb/58.5.P283] [Medline: 14507935]
23. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
24. Bevans M, Ross A, Cella D. Patient-Reported Outcomes Measurement Information System (PROMIS): efficient, standardized tools to measure self-reported health and quality of life. Nurs Outlook 2014;62(5):339-345 [FREE Full text] [doi: 10.1016/j.outlook.2014.05.009] [Medline: 25015409]
25. Revicki DA, Kawata AK, Harnam N, Chen W, Hays RD, Cella D. Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample. Qual Life Res 2009 Aug;18(6):783-791 [FREE Full text] [doi: 10.1007/s11136-009-9489-8] [Medline: 19472072]
26. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
27. Muñoz-Mendoza C, Cabrero-García J, Reig-Ferrer A, Cabañero-Martínez M. Evaluation of walking speed tests as a measurement of functional limitations in elderly people: a structured review. Int J Clin Health Psychol 2010;10(2):359-378 [FREE Full text]
28. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta-analysis. BMC Geriatr 2014 Feb 01;14:14 [FREE Full text] [doi: 10.1186/1471-2318-14-14] [Medline: 24484314]
29. Hess RJ, Brach JS, Piva SR, VanSwearingen JM. Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test. Phys Ther 2010 Jan;90(1):89-99 [FREE Full text] [doi: 10.2522/ptj.20080121] [Medline: 19959654]
30. The Python programming language for S60 mobile phones. 2010. URL: http://garage.maemo.org/projects/pys60 [accessed 2010-02-10] [WebCite Cache ID 6s2FGeUf9]
31. Ichinoseki-Sekine N, Kuwae Y, Higashi Y, Fujimoto T, Sekine M, Tamura T. Improving the accuracy of pedometer used by the elderly with the FFT algorithm. Med Sci Sports Exerc 2006 Sep;38(9):1674-1681. [doi: 10.1249/01.mss.0000227641.68360.c2] [Medline: 16960530]
32. Jette A, Haley S, Kooyoomjian J. Late-Life FDI Manual. Boston, MA: Boston University; 2002. URL: https://www.bu.edu/sph/files/2011/06/LLFDI_Manual_2006_rev.pdf [WebCite Cache ID 6s2FU762A]
33. Lachman M, Howland J, Tennstedt S, Jette A, Assmann S, Peterson E. Fear of Falling and Activity Restriction: The Survey of Activities and Fear of Falling in the Elderly (SAFE). The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 1998 Jan 01;53B(1):P43-P50. [doi: 10.1093/geronb/53B.1.P43] [Medline: 9469171]
34. Cella D, Gershon R, Bass M, Rothrock N. PROMIS Scoring Manuals. Assessment Center. 2016. URL: https://www.assessmentcenter.net/Manuals.aspx [accessed 2016-11-22] [WebCite Cache ID 6mDaRP8NU]
35. Stunkard A. A method of studying physical activity in man. Am J Clin Nutr 1960;8(5):595-601.
36. Kochersberger G, McConnell E, Kuchibhatla MN, Pieper C. The reliability, validity, and stability of a measure of physical activity in the elderly. Archives of Physical Medicine and Rehabilitation 1996 Aug;77(8):793-795. [doi: 10.1016/S0003-9993(96)90258-0] [Medline: 8702373]
37. Matthews CE, Freedson PS. Field trial of a three-dimensional activity monitor: comparison with self report. Med Sci Sports Exerc 1995 Jul;27(7):1071-1078. [Medline: 7564975]
38. Storti KL, Pettee KK, Brach JS, Talkowski JB, Richardson CR, Kriska AM. Gait speed and step-count monitor accuracy in community-dwelling older adults. Med Sci Sports Exerc 2008 Jan;40(1):59-64. [doi: 10.1249/mss.0b013e318158b504] [Medline: 18091020]
39. Rogers WA, Fisk AD. Technology design, usability, and aging: human factors techniques and considerations. In: Charness N, Warner-Schaie K, editors. Impact of Technology on Successful Aging. New York, NY: Springer Publishing; 2003:1-14.
40. Smith A. Smartphone Ownership - 2013 Update. Washington, DC: Pew Research Center; May 5, 2013.

41. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223 [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]
42. Smith AL, Chaparro BS. Smartphone text input method performance, usability, and preference with younger and older adults. Hum Factors 2015 Sep;57(6):1015-1028. [doi: 10.1177/0018720815575644] [Medline: 25850116]
43. Feuering R, Vered E, Kushnir T, Jette AM, Melzer I. Differences between self-reported and observed physical functioning in independent older adults. Disabil Rehabil 2014;36(17):1395-1401. [doi: 10.3109/09638288.2013.828786] [Medline: 24001263]
44. González-Vaca J, de la Rica-Escuín M, Silva-Iglesias M, Arjonilla-García MD, Varela-Pérez R, Oliver-Carbonell JL, et al. Frailty in INstitutionalized older adults from ALbacete. The FINAL Study: rationale, design, methodology, prevalence and attributes. Maturitas 2014 Jan;77(1):78-84. [doi: 10.1016/j.maturitas.2013.10.005] [Medline: 24189222]
45. Sharma L, Chmiel JS, Almagor O, Moisio K, Chang AH, Belisle L, et al. Knee instability and basic and advanced function decline in knee osteoarthritis. Arthritis Care Res (Hoboken) 2015 Aug;67(8):1095-1102 [FREE Full text] [doi: 10.1002/acr.22572] [Medline: 25732594]
46. Seidel UK, Gronewold J, Volsek M, Todica O, Kribben A, Bruck H, et al. Physical, cognitive and emotional factors contributing to quality of life, functional health and participation in community dwelling in chronic kidney disease. PLoS One 2014;9(3):e91176 [FREE Full text] [doi: 10.1371/journal.pone.0091176] [Medline: 24614180]
47. Shin KR, Kang Y, Oak J. [The relationship of quality of sleep, depression, late-life function and disability (LLFDI) in community-dwelling older women with urinary incontinence]. Taehan Kanho Hakhoe Chi 2008 Aug;38(4):573-581. [Medline: 18753809]
48. Lapier TK, Mizner R. Outcome measures in cardiopulmonary physical therapy: focus on the Late Life Function and Disability Instrument (LLFDI). Cardiopulm Phys Ther J 2009 Jun;20(2):32-35 [FREE Full text] [Medline: 20467536]
49. Beauchamp MK, Jette AM, Ward RE, Kurlinski LA, Kiely D, Latham NK, et al. Predictive validity and responsiveness of patient-reported and performance-based measures of function in the Boston RISE study. J Gerontol A Biol Sci Med Sci 2015 May;70(5):616-622 [FREE Full text] [doi: 10.1093/gerona/glu227] [Medline: 25512569]
50. Delbaere K, Crombez G, Van Den Noortgate N, Willems T, Cambier D. The risk of being fearful or fearless of falls in older people: an empirical validation. Disabil Rehabil 2006 Jun 30;28(12):751-756. [doi: 10.1080/09638280500304794] [Medline: 16754572]
51. Julius LM, Brach JS, Wert DM, VanSwearingen JM. Perceived effort of walking: relationship with gait, physical function and activity, fear of falling, and confidence in walking in older adults with mobility limitations. Phys Ther 2012 Oct;92(10):1268-1277 [FREE Full text] [doi: 10.2522/ptj.20110326] [Medline: 22723433]
52. Rahman S, Griffin HJ, Quinn NP, Jahanshahi M. On the nature of fear of falling in Parkinson's disease. Behav Neurol 2011;24(3):219-228 [FREE Full text] [doi: 10.3233/BEN-2011-0330] [Medline: 21876261]
53. Nilsson MH, Drake A, Hagell P. Assessment of fall-related self-efficacy and activity avoidance in people with Parkinson's disease. BMC Geriatr 2010 Oct 25;10:78 [FREE Full text] [doi: 10.1186/1471-2318-10-78] [Medline: 20973974]
54. Nilsson MH, Rehncrona S, Jarnlo G. Fear of falling and falls in people with Parkinson's disease treated with deep brain stimulation in the subthalamic nuclei. Acta Neurol Scand 2011 Jun;123(6):424-429. [doi: 10.1111/j.1600-0404.2010.01418.x] [Medline: 21492098]
55. Hornyak V, Brach JS, Wert DM, Hile E, Studenski S, Vanswearingen JM. What is the relation between fear of falling and physical activity in older adults? Arch Phys Med Rehabil 2013 Dec;94(12):2529-2534 [FREE Full text] [doi: 10.1016/j.apmr.2013.06.013] [Medline: 23816923]
56. Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. Ann Rheum Dis 2006 Nov;65 Suppl 3:iii16-iii21 [FREE Full text] [doi: 10.1136/ard.2006.059279] [Medline: 17038464]
57. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007 May;45(5 Suppl 1):S3-S11 [FREE Full text] [doi: 10.1097/01.mlr.0000258615.42478.55] [Medline: 17443116]
58. Barile JP, Reeve BB, Smith AW, Zack MM, Mitchell SA, Kobau R, et al. Monitoring population health for Healthy People 2020: evaluation of the NIH PROMIS® Global Health, CDC Healthy Days, and satisfaction with life instruments. Qual Life Res 2013 Aug;22(6):1201-1211 [FREE Full text] [doi: 10.1007/s11136-012-0246-z] [Medline: 23404737]
59. Hays RD, Spritzer KL, Thompson WW, Cella D. U.S. general population estimate for “Excellent” to “Poor” self-rated health item. J Gen Intern Med 2015 Oct;30(10):1511-1516 [FREE Full text] [doi: 10.1007/s11606-015-3290-x] [Medline: 25832617]
60. Fenton BW, Palmieri P, Diantonio G, Vongruenigen V. Application of patient-reported outcomes measurement information system to chronic pelvic pain. J Minim Invasive Gynecol 2011;18(2):189-193. [doi: 10.1016/j.jmig.2010.12.001] [Medline: 21295528]

61. Jensen RE, Potosky AL, Reeve BB, Hahn E, Cella D, Fries J, et al. Validation of the PROMIS physical function measures in a diverse US population-based cohort of cancer patients. Qual Life Res 2015 Oct;24(10):2333-2344 [FREE Full text] [doi: 10.1007/s11136-015-0992-9] [Medline: 25935353]
62. Kaiser K, Beaumont JL, Webster K, Yount SE, Wagner LI, Kuzel TM, et al. Development and validation of the functional assessment of cancer therapy-antiangiogenesis subscale. Cancer Med 2015 May;4(5):690-698 [FREE Full text] [doi: 10.1002/cam4.385] [Medline: 25619758]
63. Bingener J, Sloan JA, Seisler DK, McConico AL, Skaran PE, Farley DR, et al. PROMIS for laparoscopy. J Gastrointest Surg 2015 May;19(5):917-926 [FREE Full text] [doi: 10.1007/s11605-015-2789-0] [Medline: 25784369]
64. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. J Geriatr Phys Ther 2013;36(1):24-30. [doi: 10.1519/JPT.0b013e318248e20d] [Medline: 22415358]
65. Mathias S, Nayak US, Isaacs B. Balance in elderly patients: the “get-up and go” test. Arch Phys Med Rehabil 1986 Jun;67(6):387-389. [Medline: 3487300]
66. McDermott MM, Liu K, Ferrucci L, Criqui MH, Greenland P, Guralnik JM, et al. Physical Performance in Peripheral Arterial Disease: A Slower Rate of Decline in Patients Who Walk More. Ann Intern Med 2006 Jan 03;144(1):10-20. [doi: 10.7326/0003-4819-144-1-200601030-00005] [Medline: 16389250]
67. Sorond FA, Galica A, Serrador JM, Kiely DK, Iloputaife I, Cupples LA, et al. Cerebrovascular hemodynamics, gait, and falls in an elderly population: MOBILIZE Boston study. Neurology 2010 May 18;74(20):1627-1633 [FREE Full text] [doi: 10.1212/WNL.0b013e3181df0982] [Medline: 20479362]
68. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, et al. Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. Phys Ther 2012 Mar;92(3):388-397 [FREE Full text] [doi: 10.2522/ptj.20100398] [Medline: 22114200]
69. Balasubramanian CK. The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. J Geriatr Phys Ther 2015;38(2):78-89. [doi: 10.1519/JPT.0000000000000024] [Medline: 24949849]
70. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent systems For assessing aging changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66 Suppl 1:i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]
71. Barnes LL, Wilson RS, Bienias JL, de Leon CF, Kim HN, Buchman AS, et al. Correlates of life space in a volunteer cohort of older adults. Exp Aging Res 2007;33(1):77-93. [doi: 10.1080/03610730601006420] [Medline: 17132565]
72. Xue Q, Fried LP, Glass TA, Laffan A, Chaves PH. Life-space constriction, development of frailty, and the competing risk of mortality: the Women's Health And Aging Study I. Am J Epidemiol 2008 Jan 15;167(2):240-248. [doi: 10.1093/aje/kwm270] [Medline: 17906296]
73. Wan N, Qu W, Whittington J, Witbrodt BC, Henderson MP, Goulding EH, et al. Assessing smart phones for generating life-space indicators. Environ Plann B Plann Des 2013 Apr 01;40(2):350-361 [FREE Full text] [doi: 10.1068/b38200]

Abbreviations

ALEF: advanced lower extremity function
ANOVA: analysis of variance
BMI: body mass index
F8W: Figure-of-8 Walk
LLFDI: Late Life Function and Disability Instrument
PROMIS: Patient Reported Outcomes Measurement Information System, short form version 1.0
PROMIS-GH: Patient Reported Outcomes Measurement Information System Global Health, short form version 1.1
PROMIS-MH: Patient Reported Outcomes Measurement Information System Mental Health
PROMIS-PF: Patient Reported Outcomes Measurement Information System Physical Function
PROMIS-PH: Patient Reported Outcomes Measurement Information System Physical Health
REDCap: Research Electronic Data Capture
SAFFE: Survey of Activities and Fear of Falling in the Elderly
UEF: upper extremity function
UNMC: University of Nebraska Medical Center

Edited by C Dias; submitted 29.08.15; peer-reviewed by S Yount, C Drenowatz; comments to author 09.10.15; accepted 05.01.16; published 03.10.17.

Please cite as:
Rye Hanton C, Kwon YJ, Aung T, Whittington J, High RR, Goulding EH, Schenk AK, Bonasera SJ
Mobile Phone-Based Measures of Activity, Step Count, and Gait Speed: Results From a Study of Older Ambulatory Adults in a Naturalistic Setting
JMIR Mhealth Uhealth 2017;5(10):e104
URL: http://mhealth.jmir.org/2017/10/e104/
doi:10.2196/mhealth.5090
PMID:28974482

Original Paper

Using Mobile Phones to Improve Vaccination Uptake in 21 Low- and Middle-Income Countries: Systematic Review

Clare Oliver-Williams[1,2], PhD; Elizabeth Brown[3,4], BA (Hons); Sara Devereux[5,6], BA (Hons); Cassandra Fairhead[4,7], BA (Hons); Isaac Holeman[8,9,10], MPhil

[1]Homerton College, University of Cambridge, Cambridge, United Kingdom
[2]Cardiovascular Epidemiology Unit, Department of Public Health & Primary Care, University of Cambridge, Cambridge, United Kingdom
[3]Gonville and Caius College, University of Cambridge, Cambridge, United Kingdom
[4]University College London, London, United Kingdom
[5]Trinity College, University of Cambridge, Cambridge, United Kingdom
[6]Queens' College, University of Cambridge, Cambridge, United Kingdom
[7]King's College, University of Cambridge, Cambridge, United Kingdom
[8]Judge Business School, University of Cambridge, Cambridge, United Kingdom
[9]Global Health Academy, University of Edinburgh, Edinburgh, United Kingdom
[10]Medic Mobile, San Francisco, CA, United States

Corresponding Author:
Clare Oliver-Williams, PhD
Cardiovascular Epidemiology Unit
Department of Public Health & Primary Care
University of Cambridge
Wort's Causeway
Cambridge,
United Kingdom
Phone: 44 1223 748650
Email: cto21@medschl.cam.ac.uk

Abstract

Background: The benefits of vaccination have been comprehensively proven; however, disparities in coverage persist because of poor health system management, limited resources, and parental knowledge and attitudes. Evidence suggests that health interventions that engage local parties in communication strategies improve vaccination uptake. As mobile technology is widely used to improve health communication, mobile health (mHealth) interventions might be used to increase coverage.

Objective: The aim of this study was to conduct a systematic review of the available literature on the use of mHealth to improve vaccination in low- and middle-income countries with large numbers of unvaccinated children.

Methods: In February 2017, MEDLINE (Medical Literature Analysis and Retrieval System Online), Scopus, and Web of Science, as well as three health organization websites—Communication Initiative Network, TechNet-21, and PATH—were searched to identify mHealth intervention studies on vaccination uptake in 21 countries.

Results: Ten peer-reviewed studies and 11 studies from white or gray literature were included. Nine took place in India, three in Pakistan, two each in Malawi and Nigeria, and one each in Bangladesh, Zambia, Zimbabwe, and Kenya. Ten peer-reviewed studies and 7 white or gray studies demonstrated improved vaccination uptake after interventions, including appointment reminders, mobile phone apps, and prerecorded messages.

Conclusions: Although the potential for mHealth interventions to improve vaccination coverage seems clear, the evidence for such interventions is not. The dearth of studies in countries facing the greatest barriers to immunization impedes the prospects for evidence-based policy and practice in these settings.

(JMIR Mhealth Uhealth 2017;5(10):e148) doi:10.2196/mhealth.7792

KEYWORDS

cell phones; vaccination; communication; telemedicine; mHealth; global health

Introduction

In 2005, the World Health Organization encouraged member states to take action to incorporate eHealth in health systems and services. The term electronic health (eHealth) refers to the practice of supporting health care through information and communication technologies; eHealth initiatives have been recognized for their potential to strengthen health systems and to improve access to care [1].

The subset of eHealth initiatives that make use of mobile phones or any portable electronic devices with software applications are often discussed using the term mobile health (mHealth). Mobile technologies have been applied to a diverse range of initiatives outlined in recent reviews of mHealth interventions globally [2,3] and in low- and middle-income countries (LMIC) [4]. Given that nearly 100% of the world's population lives within reach of a mobile phone signal, many regard mHealth initiatives as particularly promising in LMIC, where other forms of communication infrastructure are underdeveloped [4]. In areas where phone ownership among the general population remains relatively low, community health workers can be key players in mHealth. Equipped with mobile phones, they can efficiently and effectively disseminate information, such as clinical updates, learning resources, and reminders, both to other health workers and to patients [5,6].

Various mHealth interventions in LMIC have aimed to improve vaccination uptake by increasing awareness of vaccine availability and providing timely reminders of when they are due. Vaccination averts approximately 2 to 3 million deaths annually and can be highly cost-effective [7]. Disparities in vaccine coverage persist because of limited resources, vaccines stock outs, geographic inaccessibility and long wait times, and poor health system management in general [8,9]. Additional demand-side barriers relate to parental knowledge and attitudes, fear of side effects, and conflicting priorities [9]. An estimated 18.7 million infants worldwide did not receive routine vaccinations such as the DPT3 (diphtheria) vaccine in 2014, and over 60% of these children live in just 10 LMIC. Evidence suggests that top-down communication strategies are detrimental to some vaccination drives in LMIC, whereas interpersonal communication incorporating local leaders and networks and utilizing a wide range of communication channels are more successful [10]. As mobile technology is widely used to improve health communication in general, mHealth interventions might be used to improve vaccine coverage.

Although the potential for mHealth interventions to improve vaccination coverage seems clear, the evidence for such interventions is not. The global population of unvaccinated children is highly concentrated in a small number of countries; as a result, literature reviews of mobile technology for immunization globally, or even of LMIC in general, may be of limited relevance. To the best of our knowledge, there has been no systematic overview of mHealth for immunization programs in countries with the greatest need to improve vaccination coverage. For this reason, the objective of this systematic review was to summarize the outcomes and implementation challenges of mHealth for vaccination interventions, focusing on 21 countries with high proportions of unvaccinated children [11].

Methods

Data Sources and Search Strategy

This systematic review was conducted using a predefined protocol and in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) and meta-analysis of observational studies in epidemiology (MOOSE) checklists (Multimedia Appendices 1 and). A literature review was conducted on February 23, 2017 (date last searched) using MEDLINE (Medical Literature Analysis and Retrieval System Online), Scopus, and Web of Science databases. Gray and white literature was also identified on the Communication Initiative Network, TechNet-21, and PATH websites. Search terms were grouped into three categories: those relating to vaccination, such as inoculation and immunization; mHealth, for example, mobile phone or telemedicine; and geographical location. No restrictions were placed on language. Details of the search terms are located in Multimedia Appendix 3. Titles and then abstracts were searched, potentially relevant papers were read, and those that did not meet the predefined inclusion criteria were removed. The inclusion criteria specified the country (Angola, Cambodia, Democratic Republic of the Congo, Ethiopia, India, Indonesia, Iraq, Kenya, Mali, Malawi, Nepal, Niger, Nigeria, Pakistan, the Philippines, Senegal, South Africa, Tanzania, Uganda, Zambia, and Zimbabwe); any form of mhealth (including mobile phone calls, phone apps, text messages, Internet, and email); and an outcome pertaining to vaccination (including uptake of vaccinations, attendance at vaccination appointments, and completeness of vaccination protocol for individuals or for regions). Reference lists of the selected studies and relevant reviews were also searched for additional publications.

Study Selection and Eligibility

Prospective interventional and observation studies that evaluated mHealth interventions on any part of a vaccination program were of interest if they were based in the relevant countries listed previously. These countries were chosen, as they include the 10 countries where more than 60% of children were unvaccinated for the final dose of Diphtheria-tetanus-pertussis vaccination as of 2014 [11], in addition to 11 countries that also have low routine vaccination uptake where the authors identified ongoing large-scale mHealth initiatives.

Data Extraction

Data were extracted by 4 authors and a predesigned data abstraction form was used. Any conflicts over inclusion were resolved by discussion. Relevant information included location, age of participants, study design, numbers included in the study, type of mobile phone intervention and frequency, duration of the study, outcome measures, and results. Where multiple publications from the same study were found, only the most up-to-date or comprehensive information was extracted.

Risk of Bias

The quality of peer-reviewed studies was rated for the risk of bias. Randomized control trials (RCTs) were assessed using the Cochrane Collaboration tool [12]. This tool considers seven different scales: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other bias. Observational studies were evaluated using the Newcastle-Ottawa Scale [13], which uses a star system to assess three aspects: participant selection, comparability of study groups, and ascertainment of outcomes. Studies that received a score of nine stars were judged to be at low risk of bias, studies that scored seven or eight stars were medium risk, and those that scored six or less were at high risk.

Analysis

Descriptive summary tables were constructed to display the results. Due to the small number and heterogeneity of studies identified, it was not possible to either conduct a statistical analysis of the results or assess publication bias through funnel plots.

Results

Studies Identified

The literature search identified 23,157 potentially relevant citations. After screening titles and abstracts, 58 peer-reviewed papers remained for further evaluation, and following detailed assessment, a further 48 were excluded (Figure 1). The remaining 10 unique papers, plus 11 studies from the gray and white literature, were included within this review.

Characteristics of Included Studies

Of the 21 studies fulfilling the inclusion or exclusion criteria, 10 were peer-reviewed, of which 3 were RCTs. Tables 1 and 2 outline the key characteristics of studies included in this review, and Table 3 summarizes geographical locations. The interventions evaluated in these papers ranged from SMS messages sent to families to remind and encourage them to take their children to the health center for vaccinations, to using mobile phones to record which settlements have been covered by vaccination campaigns, to mobile phone apps helping health workers to update and access relevant data to facilitate vaccination campaigns.

Figure 1. Flowchart for literature search.

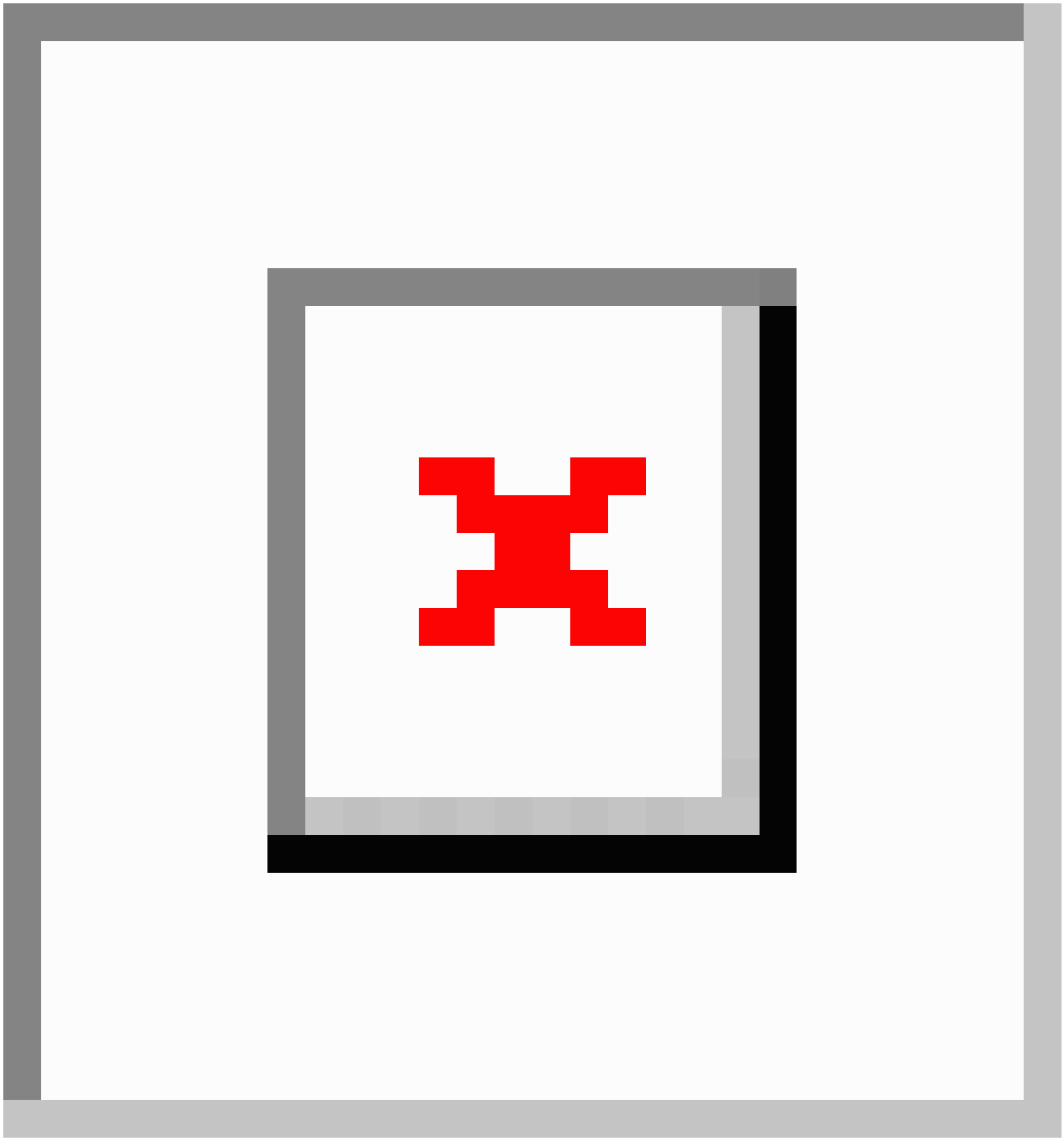


Table 1. Summary of relevant papers from peer-reviewed literature.

Lead author, date	Location	Year of study	Age range or mean age	Intervention	Outcome evaluated	Number of participants	Results
Bangure et al, 2015 [14]	Kadoma City, Zimbabwe	2013	Median age of mothers: 26 (intervention group) and 27 (controls)	SMS reminders to attend vaccination appointments at 6, 10, and 14 weeks. SMS sent 7, 3, and 1 day before appointment. Control group received routine health education only.	Percentage of children fully vaccinated with 3 doses of polio, pentavalent, and pneumococcal vaccines at 6, 10, and 14 weeks. Percentage delayed in receiving the 3 vaccines.	304 (152 intervention and 152 controls)	Vaccination coverage was greater in the intervention group (P<.001 for all): 6 weeks: 96.7% (147/152) versus 82.2% (125/152); 10 weeks: 96.1% (146/152) versus 80.3% (122/152); 14 weeks: 94.7% (144/152) versus 75.0% (114/152). Controls had a greater delay in vaccination (% delayed, median and interquartile range (IQR) delay): 6 weeks: intervention: 7.2% (11/152), 0 days (0-0), control: 76.3% (116/152), 2 days (0-6) 10 weeks: intervention: 13.2% (20/152), 0 days (0-0), control: 82.9% (126/152), 5 days (2-9) 14 weeks: intervention: 17.8% (27/152), 0 days (0-0), control: 92.1% (140/152), 10 days (6-17).
Brown et al, 2016 [15]	Ibadan, Nigeria	2012-2013	Children aged 0-12 months	Parents were randomly allocated to receive phone calls about vaccination 2 days and 1 day before the appointment, or usual care (no reminder).	Routine vaccination completion at 12 months (1 Bacillus Calmette-Guérin [BCG] dose, 4+ oral polio vaccine doses, 3 diphtheria doses, 3 hepatitis B doses, 1 measles, and yellow fever dose).	605 eligible children	The intervention group was 72% (relative risk 1.72, CI 1.50-1.98) (146 of 148 children in the intervention group vs 86 of 150 children in the control group) more likely to complete vaccination than controls who did not receive calls.
Uddin et al, 2016 [16]	Dhaka (urban) and Sunamgonj (rural), Bangladesh	2013-2014	Pregnant women, mothers with children, 0-11 months	SMS reminders sent to mothers about upcoming vaccination sessions 1 day before, at opening time, and 2 hours before closing time on the day of the vaccination.	Full vaccination rates: 1 dose of BCG; 3 doses of pentavalent (Penta) vaccine at 6, 10, and 14 weeks; and 1 dose of Measles, Mumps, and Rubella (MMR) vaccine at 9 months.	2078 children	Odds ratio (OR) for being fully vaccinated in rural areas: OR 3.6 (95% CI 1.5-8.9). OR for being fully vaccinated in urban areas: OR 2.3 (95% CI 1.1-5.5).

Lead author, date	Location	Year of study	Age range or mean age	Intervention	Outcome evaluated	Number of participants	Results
Garcia-Dia et al, 2016 [17]	Bago City area, Philippines	2013-2014	Parents of children aged 12-14 months	Participants were sent either a plain text message (SMS) or a text message with pictures once 7-10 days before the scheduled appointment date. Controls were given a verbal reminder.	MMR vaccination coverage rate Timely vaccination (difference between scheduled date of appointment visit and actual date of visit that the child was brought in for vaccination)	75 parents	Vaccination rates did not differ between the groups. Compared with verbal reminders, text reminders were associated with well-timed vaccination (difference between scheduled date of appointment and actual date of vaccination), average delay: 0.96 days for plain text reminders, 2.72 days for picture text reminders, and 20.64 days for verbal reminders (*P*=.07)
Crawford et al, 2014 [18]	Balaka District, Malawi	2011-2013	Mean age of child: 4 months	Health messages (including vaccination reminders) delivered through pushed SMS and voice messages sent to personal phones and voice messages retrieved from a community phone.	Delivery success rates for the three delivery methods User experience assessed by phone survey: Acceptability, comprehension, and self-reported behavior change.	2611 caregivers of children. A total of 1137 caregivers responded to the phone survey.	Choice of delivery system: retrieved voice messaging 63.35% (1654/2611); Pushed SMS 28.07% (733/2611); Pushed voice message 8.58% (224/2611) Delivery success: Pushed SMS 64.10% (13,053/20,363); Pushed voice 53.81% (1515/2815); Retrieved voice 27.36% (14,455/52,829). Phone survey results: 22.6% (51/226) reported not receiving any messages, most were pushed voice enrollees. 98.9% (263/266) trusted messages they received; 75.2% (200/266) recalled last message. Pushed SMS enrollees were more likely to report intended or actual change in behavior (91%, 87/96) than pushed (56%, 17/30,) or retrieved (65.7%, 92/140,) voice enrollees; *P*=.01.

Lead author, date	Location	Year of study	Age range or mean age	Intervention	Outcome evaluated	Number of participants	Results
Kazi et al, 2014 [19]	Karachi, Pakistan	2012-2013	Not given (NG)	SMS messages sent to caregivers to monitor coverage of polio supplementary immunization activities (SIAs): 1. Did the vaccinator visit your home? 2. Did [child] receive polio vaccine? US $0.20 of phone credit was given for replying. Nonresponders were contacted via direct phone calls.	Proportion of caregivers who replied to the SMS or follow-up phone calls. Estimates of vaccine coverage achieved during polio SIAs obtained through automated SMS and currently used methods for estimating vaccine coverage, as utilized by the World Health Organization.	Across 7 districts, 5880 randomly sampled caregivers of a child <5 years	Response rate: first SMS 22.99% (1352/5880); second SMS 14.00% (823/5880). 74.90% (4404/5880) of participants did not respond to SMS messages, of whom, 56.00% (2466/4404) responded to an investigator's phone calls. Those who responded to calls had similar levels of vaccine coverage to those who responded to SMSs. Reasons given for not responding to SMS (of caregivers who were contactable by direct phone call): "Too busy" 36.01% (888/2466); "Not interested" 32.00% (789/2466); "Unable to read the message" 20.00% (493/2466).
Wakadha et al, 2013 [20]	30 villages within 5 km of Ting'Wan'I hospital in Western Kenya.	2011	Mothers of children up to 4 weeks of age at baseline	Reminder SMS sent (3 days prior and on day of vaccination) for 2 doses of pentavalent vaccination. If the child was vaccinated on time, the mother was given approximately US $2. If the child was not vaccinated, another reminder was sent.	(1) Percentage of children vaccinated at hospital or other health facilities. (2) Percentage who did not receive SMSs (3) Percentage with mobile phone access 4) Follow-up at 14 weeks: influence of financial reward on vaccination	72 mothers (first dose: 69 sent SMS reminders, 3 not sent, second dose: 44 sent SMS)	(1) First dose: 70% (48/69) vaccinated at Ting'wang'I, Hospital, 10% (7/69) at other hospitals. Second dose: 91% (40/44) vaccinated at TWI hospital, 5% (2/44) at other hospitals. (2) Of the 38% (27/72) not sent SMS, 26% (7/27) vaccinated at TWI, 19% (5/27) at other hospitals, 30% (8/27) not vaccinated, and 26% (7/27) unknown. (3) 26% (19/72) had their own phone, and 74% (53/72) had access to another person's phone (4) Forty-nine mothers reported reminders influenced their decision to vaccinate.
Touray et al, 2016 [21]	10 states in northern Nigeria	2012-2015	NG	Global positioning system–enabled Android phones were given to vaccination teams and were used to record team tracks.	Settlements covered by vaccination teams during polio campaigns	NG	There was a reduction in chronically missed settlements (those missed in the last 3 campaigns): 2014—5833 settlements, 2015—1257 settlements. There was an increase in the number of missed settlements: 2014—4142, 2015—7008.

Lead author, date	Location	Year of study	Age range or mean age	Intervention	Outcome evaluated	Number of participants	Results
Balakrishnan et al, 2016 [22]	Bihar, India	2012-2014	NG	Mobile-based tool for health workers that registers when vaccinations are due and administered, creating electronic records.	Received 1+ tetanus vaccine	512 frontline workers, 19,888 children registered	Coverage in implementation area (95% CI): 79.38% (58.90-80.26) (15,771 children vaccinated of 19,888 registered) Coverage in implementation area in the previous year (%): 74.12 Coverage in rest of Bihar (%): 80
Mbabazi et al, 2015 [23]	Kenya: 8 districts of Nairobi and 3 from Nyanza or western provinces	2012	Children aged 9-59 months	A Web-enabled mobile phone app recording house visits (3 days prior and 4 days after vaccination campaigns), vaccinations, and relaying information to campaign organizers.	Percentage of households aware of the campaign before start; Percentage planning to vaccinate their children Post campaign: Percentage of households with children vaccinated against measles; Percentage with a confirmed vaccination.	164,643 houses (161,695 children) pre campaign; 175,617 houses (180,493 children) post campaign	56.00% (92,200/164,643) of households had heard about the campaign. 75.00% (123,482/164,643) of households planned to bring their children for vaccination. 96.00% (168,592/175,617) of households reported children having had a measles vaccination post campaign, and 92.00% (161,568/175,617) of households had children with a confirmed vaccination.

Eleven initiatives were identified from gray and white literature; eight took place in India, two in Pakistan, and one in Zambia. Several programs involved more than one intervention, including messages sent to parents to encourage their children to get vaccinated, information about vaccination made freely accessible via mobile phone, tools to identify unvaccinated children with the health authority using SMS, data management tools for health workers (such as electronic vaccination records and a mobile phone app to track where vaccinations have been administered and control supplies), and tools to help health workers persuade hesitant families.

SMS Reminders for Vaccinations

Eight peer-reviewed studies reported the use of phone calls or SMS reminders for vaccinations, two of which additionally offered cash incentives. The three studies that did not offer cash incentives included an RCT by Bangure et al [14] conducted in Zimbabwe. SMS reminders were sent to parents (n=152) when their baby was 6, 10, and 14 weeks old, in addition to routine health education. The control group received health education alone (n=152). At all three time points, the percentage of children fully vaccinated with the relevant dose of polio, pentavalent, and pneumococcal vaccines was significantly higher in the intervention than the control group (<.001), and the delay in receiving the vaccinations was significantly less in the intervention than the control group (<.001). Another RCT by Brown et al [15] conducted in Nigeria identified increased coverage rates relative to the usual care when receiving phone call reminders 2 days and 1 day before a vaccination appointment (Relative risk 1.72, 95% CI 1.50-1.98). Uddin et al [16] similarly found that SMS reminders increased the odds of vaccination uptake in both urban and rural areas; odds ratio (OR) 2.3 (95% CI 1.1-5.5) and OR 3.6 (95% CI 1.5-8.9), respectively. Garcia-Dia et al [17] assessed coverage rates in an RCT in the Philippines after 75 parents were sent either a plain text message (short service message, SMS), a text message with pictures, or a verbal reminder. Although vaccination rates did not differ by reminder, text reminders with and without a picture were associated with a shorter delay in receiving the vaccination than verbal reminders. Crawford et al [18] sent SMS or voice messages to either the personal or community phone of 2611 caregivers of children under the age of 1 year. Pushed SMS messaging (where a message is sent to a phone's notification center or status bar) was the most successful mode of delivery (64.10%, 13,053/20,363, of sent messages were received). However, most women did not own a mobile phone, so similar numbers of messages were delivered by retrieved voicemail to community phones and pushed SMS to personal phones. No control group was included, but the majority of individuals who received messages trusted (98.8%; 263/266) and could recall (72.2%; 200/277) those messages.

Table 2. Summary of studies from white and gray literature.

Name of study or source of participants	Location	Year of study	Age range or mean age	Intervention form	Intervention period and regularity of intervention	Outcome evaluated	Total number of participants	Results
MIRA channel [24]	Haryana, India	2012-ongoing	Children	Integrated mobile phone channel with health information to women and connecting them with health services.	Continuous	Vaccination rates	Not given (NG)	Increase in vaccination rates by 41% (from 51% to 92%, overall rates in Haryana: 78%).
Mobile Kunji [25]	Bihar, India	2011-2015	NG	When a health worker dials the number, they can play a health message—voiced by a character called Dr Anita, an engaging but authoritative female doctor—to the family via their mobile phone.	NG	Percentage of children unvaccinated, Percentage of children (6-11 months) receiving DPT2 (diphtheria) vaccine, Percentage of children (<11 months) with a vaccination card	NG	Mobile Kunji was not found to significantly alter vaccination uptake [26].
UNICEF, India's National Immunization Day [27]	India	1999-2000	Children	India's national telecom authority agreed to replace the ringtone with a recorded message reminding the public about the date of the National Immunization Day.	Annually	Number of children vaccinated for polio, Percentage of coverage (2+ doses), Percentage of zero doses, number of polio cases	NG	151 million children vaccinated, 98.6% coverage (at least 2 doses), 0.7% of children with zero doses, 265 cases of polio in 2000.
Mobilink [28,29]	Pakistan	2009-2012	Children	Subscribers to the Mobilink mobile operator will be able to report areas and children where the polio vaccination teams have not reached. The respective health authority will then be in contact to vaccinate the missed children. Mobilink also sends an SMS to create awareness about polio.	Period: 1-3 days, with >3 rounds for reporting unvaccinated children.	NG	NG	15,000 SMS messages about unvaccinated children were received during February 15-17, 2010.
Aarogyam [30]	Uttar Pradesh, India	2008 onwards	Children under 5 years	Health alerts are sent to parents about vaccination through an SMS and phone calls.	NG	Vaccination coverage	NG	Vaccination coverage has shown a significant positive trend over time. Polio, Bacillus Calmette-Guérin (BCG), measles, and tetanus coverage has gone up from approximately 60% in 2008 to 91% in 2010.

Name of study or source of participants	Location	Year of study	Age range or mean age	Intervention form	Intervention period and regularity of intervention	Outcome evaluated	Total number of participants	Results
Khushi Baby [31]	Northern India	2015	Infants	Electronic copy of the vaccination record stored on a necklace. Health workers scan the necklace using an app on their mobile phone to transfer vaccination data to the necklace. Data are also automatically uploaded to "the cloud." Parents get vaccination reminder voice calls.	Continuous	NG	NG	Pilot study is ongoing.
mSakhi [32]	Uttar Pradesh, India	2011	NG	A mobile-based interactive multimedia learning app for health workers.	4 months	Increase in knowledge in maternal-newborn care (including vaccination)	25 health visitors	Qualitative data indicated improved counseling during home visits and increased credibility of health workers in the community.
HealthPhone [33]	India	2009-ongoing	Children	Video reference library that covers vaccination and SMS messages for those who cannot access video.	No time limit, continuous	Multiple health outcomes, including uptake of vaccines	NG	"After we put HealthPhone into the hands of village women...their health and the health of their children dramatically improved."
Freedom Polio [34]	India	2012	Children under 5 years	An app that allows health works to track where polio vaccinations have been administered.	No time limit, continuous	NG	21 million children	NG
UNICEF, Zambian Health Ministry, two mobile phone companies, Zain and Mobile Telephone Networks [35]	28 districts in Zambia	2009	Children under 5 years	SMS: "Your child can be healthier! Take your children under age five to the nearest health centre for free vaccinations from 20-25 July."	NG	NG	NG	NG
Interactive Research and Development's (IRD) Interactive alerts [36,37]	Karachi, Pakistan	2012 onwards	NG	Mobile phone–based vaccine registry system that uses SMS reminders to caregivers and conditional cash transfers to caregivers and health workers.	NG	Vaccination coverage and timeliness	14,000 infants	Interim data analysis suggests improved immunization coverage and timeliness; an impact evaluation study is underway to assess this more thoroughly.

Table 3. Geographical locations of the included studies.

Location	Number of studies included
Bangladesh	1
India	9
Kenya	2
Malawi	1
Nigeria	2
Pakistan	3
The Philippines	1
Zambia	1
Zimbabwe	1

Cash Incentives to Increase Vaccination Uptake

Two studies used cash incentives to increase vaccination uptake while sending SMS reminders. Kazi et al [19] used SMS messages sent to 5880 caregivers in Pakistan, along with a conditional cash transfer in the form of approximately US $0.20 of phone credit, to monitor polio vaccination coverage. Response rates to the SMS messages were low (74.90%, 4404/5880, of participants did not respond). The initial nonresponders who were followed up by phone call had similar rates of vaccination uptake to those who responded to the SMS messages. Wakadha et al [20] conducted a pilot study exploring the feasibility of setting up an integrated mobile phone–based system to remind and incentivize mothers (n=72) to vaccinate their children in rural Kenya. Mothers received SMS reminders of vaccination dates and conditional cash transfers of either mPESA (a mobile-based money transfer service) credit or phone credit, if the child was vaccinated within 4 weeks of the scheduled date. The small sample size and lack of a comparison group meant that it was not possible to draw conclusions about the program's effectiveness, but enrolled mothers reported mostly positive experiences at the end of the study, and most mothers did have access to a phone. Importantly, this study was limited by its focus on a single facility. Caregivers who took children to nearby facilities for vaccinations were recorded as unvaccinated by the first facility and thereafter, were not sent additional SMS.

Mobile-Based Interactive Apps for Health Workers

Three studies used mobile-based interactive learning apps to aid or track the progress of health workers in vaccination. Touray et al [21] utilized the global positioning system of Android phones to track where vaccination teams had been, which helped reduce the number of settlements in northern Nigeria that were not covered in the last three campaigns from 5833 in 2014 to 1257 in 2015. Balakrishnan et al [22] found no improvement in coverage for tetanus when health workers used a mobile phone tool that created electronic vaccination records and registered when vaccination was due and administered. Mbabazi et al [23] evaluated a mobile phone app used by health workers that was designed to assess the awareness and intention to take part in a measles vaccination campaign before the campaign's onset at house visits, as well as to evaluate the uptake of the vaccinations after the campaign. Of the more than 150,000 households included in the survey, approximately half were aware of the vaccination campaign, and once informed, 74.99% (123,482/164, 643) of households planned to bring their children in for vaccination. After the campaign, 95.99% (168,592/175,617) of households reported their child had received a measles vaccination, and 92.00% (161, 568 of 175,617) had this independently confirmed. This intervention was found to reduce misconceptions about vaccination, and the use of the mobile phone app to assess uptake of vaccination helped inform service delivery plans.

Risk of Bias

Of the 10 peer-reviewed studies, three studies did not evaluate controls or individuals unexposed to the intervention, so it was not possible to evaluate their risk of bias using the Newcastle-Ottawa Scale or the Cochrane Collaboration tool. In the eight studies that could be evaluated, the risk of selection bias affecting the results was judged to be low for one RCT by Bangure et al but with a higher risk of bias in the other two RCTs (Table 4). The risk of bias in the observational studies was also deemed to be high, with most of the concern regarding the possibility of outcome bias and bias arising from a lack of comparability (Table 5).

Table 4. Assessment of bias in randomized controlled trials.

Paper	Risk of bias						
	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other bias
Bangure et al [14]	Low	Low	Unclear	Unclear	Low	Low	Low
Brown et al [15]	Low	Low	High	Medium	Low	Low	Low
Garcia-Dia et al [17]	Low	Medium	High	Unclear	Low	Low	Low

Table 5. Assessment of bias in observational cohort studies.

Paper	Selection[a]	Comparability[b]	Outcome or exposure[c]
Uddin et al [16]	2	2	2
Wakadha et al [20]	4	0	1
Touray et al [21]	3	0	3
Balakrishnan et al [22]	2	0	1

[a]Maximum score is 4.

[b]Maximum score is 2.

[c]Maximum score is 3.

Initiatives Identified From White and Gray Literature

Of the eleven initiatives identified, six showed some evidence of impact on vaccination rates. Implementation of the MIRA channel [24] (an integrated mobile phone channel providing health information to women and connecting them with health services) corresponded with a 41% increase in vaccination rates. A program using the Mobile Kunji program [25,26] (in which the health worker can play a health message to the family via their mobile phone) recorded a 5% decrease in the percentage of children (6-11 months) unvaccinated with the first diphtheria vaccine and a 6% increase in children receiving the second diphtheria vaccine. Another successful strategy included the involvement of India's national telecom authority who replaced the dial tone on mobile phones with a recorded message that reminded the public of National Immunization Day [27], whereas the Mobilink mobile operator in Pakistan recorded 13,000 SMS messages about unvaccinated children during the annual polio vaccination campaign in 2010 and circulated seven million SMS reminders in 2009 [28,29]. Aarogyam [30] reported improved vaccination uptake through the use of automatic voice calls and SMS reminders sent to parents about vaccination appointments, among other postnatal care.

Unfortunately, other initiatives did not provide quantitative results. Some pilot studies are ongoing (Khushi Baby [31]), and one provided qualitative evidence of improved knowledge [32]. No outcomes were found for three studies, one study used educational videos accessible on mobile phones [33], another looked at an app to track vaccinations [34], and two studies used SMS vaccination reminders [35-37].

Discussion

Overall, mHealth technology can and has been used to increase vaccination uptake in LMIC, but the quality of the evidence is limited, and further research is needed to better quantify its potential impact and to determine the most effective strategies.

Evidence That mHealth Interventions Can be Effective in Increasing Vaccination Uptake

The literature reviewed indicates that mobile technology can be used in a variety of ways to improve vaccination uptake. Although most studies lacked comparison groups, the results broadly suggest an improved uptake of vaccinations with mobile phone–based interventions.

SMS reminders for vaccination appointments were found to increase uptake and reduce delays in receiving vaccinations in Zimbabwe [14]. In Kenya, mothers who received SMS reminders about vaccination appointments reported mostly positive experiences [23]. A decrease in the percentage of unvaccinated children and an increase in the number of children with a vaccination card were found when health care workers used their mobile phones to play a prerecorded message to families. Furthermore, a 41% increase in vaccination rates was observed in rural India after the introduction of an integrated mobile channel providing health information and connecting mothers with health services [24].

However, some studies reported no improvement upon intervention. A study from Pakistan found low response rates to SMS messages about vaccinations, even when a financial reward was attached [19].

Challenges in the Use of mHealth Interventions to Increase Vaccination Uptake

The studies we reviewed, as well as the related research that these studies cited to explain the design of their interventions, raise a number of challenges that can impede the integration of mobile phones into vaccination programs. Several of these studies discuss rates of phone ownership in their particular areas of intervention, including how these differ among men and women (Crawford et al, Uddin et al, and Kazi et al). In LMIC, generally, women are 21% less likely to own a mobile phone than men (increasing to 37% in Asia) [38]. As women are the primary caregivers to children, this may impact mHealth vaccination interventions. Furthermore, two-thirds of illiterate adults are women [39], which can further reduce the effectiveness of SMS messages. In households where the father owns the mobile phone, it is imperative that the father is engaged in the project, as exemplified in the study by Wakadha et al [20], where in a few cases husbands did not approve of the study.

Frequent exposure to SMS messages can result in the effectiveness of the message being weakened; in a different setting, Strandbygaard et al found that participants stopped reading reminder messages after a few weeks [40]. Therefore, the effectiveness of messages of different length and over time needs to be assessed when sending SMSs.

Developing the appropriate infrastructure [9] and ensuring adequate resources are available is important. Weaknesses in other areas of the health system may render mHealth interventions aiming to increase demand for services meaningless: mHealth can improve access to vaccines only as long as they remain consistently available from health centers [9]. Additionally, increasing demand for vaccination can have unintended consequences. One study reported that extensive and comprehensive communication campaigns for 15 new vaccines led to greater demand for vaccination in a number of LMIC. However, high demand resulted in vaccine shortages, which later thwarted the increased demand [41]. For this reason, policy makers and implementers of mHealth interventions to improve vaccination programs should be aware that eHealth interventions in general [42], and mHealth interventions in particular [6,43] are deeply complex and context-dependent.

Limitations, Opportunities, and Need for Further Research

Of the literature reviewed, the included studies were predominantly observational studies that appraised process and usage output. These had various methodological limitations such as (1) sample selection based on convenience, without randomization; (2) small sample sizes; (3) lack of information on process validation, including recruitment type, response rate, and retention rate; and (4) no control groups. These limitations make the conclusions of the observational studies less secure. Given the potential of mHealth, RCTs in LMIC to determine the efficacy of using mHealth for vaccinations are needed.

It is clear that there are a number of gaps in the literature concerning this topic in the countries of interest. Relatively more compelling evidence exists for mHealth interventions addressing demand-side barriers to service uptake, whereas fewer evaluated interventions aim to boost immunization by strengthening health systems through data management, decision support, or provider training and education. This is a notable gap because reviews of why children go unvaccinated document not only highlight gaps in household knowledge and attitudes but also issues related to poor service quality and accessibility [8,9]. Moreover, there is reason to believe that this gap can be addressed because outside immunization programs, mHealth interventions have been widely used as strengthening tools for health systems [4]. For example, given that a study on stock tracking of malaria medications in Tanzania showed a 52% reduction in medication stock-outs within 21 weeks of the induction of weekly SMS requesting stock counts [44], there is a precedent for the integration of mHealth into vaccination stock control.

Conclusions

There is reason to be optimistic regarding the potential for mobile phones to increase vaccination coverage in LMIC. Mobile technologies are flexible and widely available tools that can be utilized in myriad ways. This review provides evidence of potential effectiveness for SMS reminders to families regarding vaccination, as well as for educational tools for health workers.

However, the research is preliminary and limited. Further research is needed to determine the most effective mHealth interventions and to refine their use, for example, clarifying the optimal schedule of reminders for programs using SMS reminders of vaccination appointments. It will also be necessary to evaluate different mHealth interventions against each other, and against other potential programs, to examine their comparative cost-effectiveness at increasing vaccination coverage. mHealth interventions addressing vaccination stock-outs, cold storage, or other health systems strengthening challenges merit further study.

Overall, there is preliminary evidence to support the use of mHealth technology to increase vaccination coverage in LMIC. However, further research is needed to guide and improve the use of these technologies in the future and to strengthen the case for their cost-effectiveness.

Acknowledgments

CO screened titles and abstracts, constructed tables, and contributed to writing and amending of the manuscript. EB, SD, and CF screened titles and abstracts and contributed to writing and amending of the report. IH developed the concept and contributed to writing and amending of the report. The study was funded by the Medical Research Council, British Heart Foundation, and Homerton College, Cambridge. The nonprofit organization Medic Mobile helped to clarify the need for this study and helped the

study authors to identify relevant gray literature. The student-run think tank Polygeia set up the commission with Medic Mobile and supported the authors.

Conflicts of Interest

IH is a board member at Medic Mobile, for which he receives no compensation. Medic Mobile is a nonprofit organization that equips health workers with free and open source digital tools to strengthen global health programs. All authors declare no competing financial interests.

Multimedia Appendix 1

PRISMA 2009 checklist.

[PDF File (Adobe PDF File), 117KB - mhealth_v5i10e148_app1.pdf]

Multimedia Appendix 2

MOOSE checklist.

[PDF File (Adobe PDF File), 73KB - mhealth_v5i10e148_app2.pdf]

Multimedia Appendix 3

Search terms used to identify relevant published literature.

[PDF File (Adobe PDF File), 31KB - mhealth_v5i10e148_app3.pdf]

References

1. Apps.who. 58th World Health Assembly URL: http://apps.who.int/gb/ebwha/pdf_files/WHA58-REC1/english/A58_2005_REC1-en.pdf [WebCite Cache ID 6oRxo3M2W]
2. 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]
3. 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 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
4. 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]
5. 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]
6. Holeman I, Evans J, Kane D, Grant L, Pagliari C, Weller D. Mobile health for cancer in low to middle income countries: priorities for research and development. Eur J Cancer Care (Engl) 2014 Nov;23(6):750-756. [doi: 10.1111/ecc.12250] [Medline: 25324023]
7. WHO. Health Topics: Immunization URL: http://www.who.int/topics/immunization/en/ [accessed 2017-03-30] [WebCite Cache ID 6pLl9zCOR]
8. WHO. Immunization coverage fact sheet URL: http://www.who.int/mediacentre/factsheets/fs378/en/ [WebCite Cache ID 6oSlKG3Nb]
9. Mechael P, Batavia H, Kaonga N, Searle S, Kwan A, Goldberger A, et al. Globalproblems-globalsolutions. Barriers and gaps affecting mHealth in low and middle income countries: policy white paper URL: http://www.globalproblems-globalsolutions-files.org/pdfs/mHealth_Barriers_White_Paper.pdf [accessed 2017-02-22] [WebCite Cache ID 6oSk7TPSa]
10. Comminit. Lessons learned from five country studies of communication support for polio eradication and routine URL: http://comminit.com/hiv-aids/content/lessons-learned-five-country-studies-communication-support-polio-eradication-and-routine [WebCite Cache ID 6oS7OuvdD]
11. WHO. Global immunization data URL: http://www.who.int/immunization/monitoring_surveillance/global_immunization_data.pdf [accessed 2017-03-30] [WebCite Cache ID 6pLlEWnBw]
12. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011;343:d5928 [FREE Full text] [Medline: 22008217]

13. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. Ohri. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [WebCite Cache ID 6oS7kdgJv]
14. Bangure D, Chirundu D, Gombe N, Marufu T, Mandozana G, Tshimanga M, et al. Effectiveness of SMS reminders on childhood immunization programme in Kadoma, Zimbabwe - a randomized controlled trial, 2013. BMC Public Health 2015 Feb 12;15:137 [FREE Full text] [doi: 10.1186/s12889-015-1470-6] [Medline: 25885862]
15. Brown VB, Oluwatosin OA, Akinyemi JO, Adeyemo AA. Effects of community health nurse-led intervention on childhood routine immunization completion in primary health care centers in Ibadan, Nigeria. J Community Health 2016 Apr;41(2):265-273. [doi: 10.1007/s10900-015-0092-3] [Medline: 26395786]
16. Uddin MJ, Shamsuzzaman M, Horng L, Labrique A, Vasudevan L, Zeller K, et al. Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh. Vaccine 2016 Jan 4;34(2):276-283. [doi: 10.1016/j.vaccine.2015.11.024] [Medline: 26647290]
17. Garcia-Dia MJ, Fitzpatrick JJ, Madigan EA, Peabody JW. Using text reminder to improve childhood immunization adherence in the Philippines. Comput Inform Nurs 2017;35(4):212-218. [doi: 10.1097/CIN.0000000000000307] [Medline: 27828815]
18. Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. Glob Health Sci Pract 2014;2(1):35-46 [FREE Full text] [doi: 10.9745/GHSP-D-13-00155] [Medline: 25276561]
19. Kazi AM, Murtaza A, Khoja S, Zaidi AK, Ali SA. Monitoring polio supplementary immunization activities using an automated short text messaging system in Karachi, Pakistan. Bull World Health Organ 2014 Mar 1;92(3):220-225 [FREE Full text] [doi: 10.2471/BLT.13.122564] [Medline: 24700982]
20. Wakadha H, Chandir S, Were EV, Rubin A, Obor D, Levine OS, et al. The feasibility of using mobile-phone based SMS reminders and conditional cash transfers to improve timely immunization in rural Kenya. Vaccine 2013 Jan 30;31(6):987-993 [FREE Full text] [doi: 10.1016/j.vaccine.2012.11.093] [Medline: 23246258]
21. Touray K, Mkanda P, Tegegn S, Nsubuga P, Erbeto TB, Banda R, et al. Tracking vaccination teams during polio campaigns in northern Nigeria by use of geographic information system technology. J Infect Dis 2016;213(Suppl 3):S67-S72. [Medline: 26609004]
22. Balakrishnan R, Gopichandran V, Chaturvedi S, Chatterjee R, Mahapatra T, Chaudhuri I. Continuum of care services for maternal and child health using mobile technology - a health system strengthening strategy in low and middle income countries. BMC Med Inform Decis Mak 2016 Jul 07;16:84 [FREE Full text] [doi: 10.1186/s12911-016-0326-z] [Medline: 27387548]
23. Mbabazi WB, Tabu CW, Chemirmir C, Kisia J, Ali N, Corkum MG, et al. Innovations in communication technologies for measles supplemental immunization activities: lessons from Kenya measles vaccination campaign, November 2012. Health Policy Plan 2015 Jun;30(5):638-644 [FREE Full text] [doi: 10.1093/heapol/czu042] [Medline: 24920218]
24. Mirachannel. MIRA channel - women mobile lifeline channel URL: http://www.mirachannel.org/ [accessed 2017-02-21] [WebCite Cache ID 6oS7yERNm]
25. Rethink1000days. Mobile kunji?: shaping demand and practices URL: https://www.rethink1000days.org/programme-outputs/mobile-kunji/ [accessed 2017-02-21] [WebCite Cache ID 6oS8jWvQU]
26. Chamberlain S. Mitpressjournals. A mobile guide toward better health: how mobile kunji is improving birth outcomes in Bihar, India URL: http://mitpressjournals.org/userimages/ContentEditor/1415302178306/INNOVATIONS_DIGITAL-INCLUSION.pdf [accessed 2017-03-29] [WebCite Cache ID 6pKQA7m9s]
27. WHO. The use of mobile phones in polio eradication URL: http://www.who.int/bulletin/volumes/94/2/15-163683/en/ [accessed 2017-02-23] [WebCite Cache ID 6oUUSV7c2]
28. Pktelecomnews.blogspot. Mobilink concludes awareness drive URL: http://pktelecomnews.blogspot.co.uk/2010/07/mobilink-concludes-awareness-drive.html [accessed 2017-02-23] [WebCite Cache ID 6oUPJTtzD]
29. Thefreelibrary. Seven million SMS disseminated to boost anti-polio drive URL: https://www.thefreelibrary.com/403.htm [accessed 2017-02-23] [WebCite Cache ID 6oUOovIyK]
30. Indiagovernance. Aarogyam URL: http://indiagovernance.gov.in/files/gkc_oneworld_aarogyam.pdf [accessed 2017-03-29] [WebCite Cache ID 6pKQiCgRj]
31. Khushibaby. URL: http://www.khushibaby.org/ [accessed 2017-02-21] [WebCite Cache ID 6oS9Yk92T]
32. Kumar A, Bora G, Jana L. Papers.eletsonline. mSakhi putting technology into the hands of community health workers URL: http://papers.eletsonline.com/2011/11/21/msakhi-putting-technology-in-the-hands-of-community-health-workers/ [accessed 2017-03-30] [WebCite Cache ID 6pLmW0v4e]
33. YouTube. HealthPhoneTM: how immunization protects: completing the immunization cycle protects children URL: https://www.youtube.com/watch?v=Wr-RKyifFmI [accessed 2017-02-23] [WebCite Cache ID 6oUX74xdN]
34. Quraishi S. Freedompolio. Freedom Polio URL: http://www.freedompolio.org/ZMQ_FreedomPolio_Brochure.pdf [accessed 2017-02-23] [WebCite Cache ID 6oUM9Cf3I]
35. UN News Centre. UN. UNICEF uses text messages to spread the word about polio in Zambia URL: http://www.un.org/apps/news/story.asp?NewsID=31517#.WK1U1RKLQcg [WebCite Cache ID 6oSkkb0sA]

36. Chandir S, Zaidi A, Khowaja S, Gonzales F. Africanstrategies4health. Interactive alerts: improving vaccine coverage through small incentives URL: http://www.africanstrategies4health.org/uploads/1/3/5/3/13538666/interactive_alerts.pdf [WebCite Cache ID 6pKQv0DGw]
37. Zaidi A, Khowaja S. Apps.WHO. Small incentives improve vaccine coverage in Pakistan URL: http://apps.who.int/iris/bitstream/10665/92806/1/WHO_RHR_13.16_eng.pdf [WebCite Cache ID 6pKR4YFBo]
38. Cherieblairfoundation. Women and mobile: a global opportunity URL: http://www.cherieblairfoundation.org/women-and-mobile-a-global-opportunity/ [accessed 2017-07-11] [WebCite Cache ID 6rsUsDuzD]
39. UNESCO. Literacy URL: http://www.unesco.org/new/en/education/themes/education-building-blocks/literacy/resources/statistics [WebCite Cache ID 6oS8EEgiE]
40. 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 [FREE Full text] [doi: 10.1016/j.rmed.2009.10.003] [Medline: 19854632]
41. Comminit. Bottlenecks and breakthroughs: lessons learned from new vaccine introductions in low-resource countries, 2008 to 2013 URL: http://www.comminit.com/global/content/bottlenecks-and-breakthroughs-lessons-learned-new-vaccine-introductions-low-resource-cou [WebCite Cache ID 6oS8WxE5A]
42. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
43. PLOS Medicine Editors. A reality checkpoint for mobile health: three challenges to overcome. PLoS Med 2013;10(2):e1001395 [FREE Full text] [doi: 10.1371/journal.pmed.1001395] [Medline: 23468597]
44. Barrington J, Wereko-Brobby O, Ward P, Mwafongo W, Kungulwe S. SMS for Life: a pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology. Malar J 2010 Oct 27;9:298 [FREE Full text] [doi: 10.1186/1475-2875-9-298] [Medline: 20979633]

Abbreviations

eHealth: electronic health
LMIC: low- and middle-income countries
mHealth: mobile health
OR: odds ratio
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 30.03.17; peer-reviewed by J Crawford, P Kawale; comments to author 14.06.17; revised version received 11.07.17; accepted 11.07.17; published 04.10.17.

Please cite as:
Oliver-Williams C, Brown E, Devereux S, Fairhead C, Holeman I
Using Mobile Phones to Improve Vaccination Uptake in 21 Low- and Middle-Income Countries: Systematic Review
JMIR Mhealth Uhealth 2017;5(10):e148
URL: http://mhealth.jmir.org/2017/10/e148/
doi:10.2196/mhealth.7792
PMID:28978495

Viewpoint

Tackling Regional Public Health Issues Using Mobile Health Technology: Event Report of an mHealth Hackathon in Thailand

Atipong Pathanasethpong[1], MSc, MD; Chitsutha Soomlek[2], PhD; Katharine Morley[3], MPH, MD; Michael Morley[3], ScM, MD; Pattarawit Polpinit[4], PhD; Alon Dagan[5], MD; James W Weis[6], SM; Leo Anthony Celi[6], MPH, MSc, MD

[1]Department of Anesthesiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand

[2]Department of Computer Science, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand

[3]Harvard Medical School, Boston, MA, United States

[4]Department of Computer Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, Thailand

[5]Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Boston, MA, United States

[6]Massachusetts Institute of Technology, Cambridge, MA, United States

Corresponding Author:
Atipong Pathanasethpong, MSc, MD
Department of Anesthesiology
Faculty of Medicine
Khon Kaen University
123 Mittraphab Road
Khon Kaen,
Thailand
Phone: 66 891758278
Email: atipat@kku.ac.th

Abstract

Hackathons are intense, short, collaborative events focusing on solving real world problems through interdisciplinary teams. This is a report of the mHealth hackathon hosted by Khon Kaen University in collaboration with MIT Sana and faculty members from Harvard Medical School with the aim to improve health care delivery in the Northeast region of Thailand. Key health challenges, such as improving population health literacy, tracking disease trajectory and outcomes among rural communities, and supporting the workflow of overburdened frontline providers, were addressed using mHealth. Many modifications from the usual format of hackathon were made to tailor the event to the local context and culture, such as the process of recruiting participants and how teams were matched and formed. These modifications serve as good learning points for hosting future hackathons. There are also many lessons learned about how to achieve a fruitful collaboration despite cultural barriers, how to best provide mentorship to the participants, how to instill in the participants a sense of mission, and how to match the participants in a fair and efficient manner. This event showcases how interdisciplinary collaboration can produce results that are unattainable by any discipline alone and demonstrates that innovations are the fruits of collective wisdom of people from different fields of expertise who work together toward the same goals.

(JMIR Mhealth Uhealth 2017;5(10):e155) doi:10.2196/mhealth.8259

KEYWORDS

hackathon; mHealth; interdisciplinary collaboration

Introduction

Khon Kaen University and its Role in the Health Care of Northeast Thailand

Health care for the population of Thailand is covered under a universal coverage scheme subsidized by the government [1]. Every citizen is entitled to free health care accessed by first visiting a local primary health center. Patients requiring more advanced capabilities are referred to more sophisticated facilities. In this system, university-based health centers serve as the pinnacle of advanced care and are the ultimate referral destinations for patients with complex diseases and conditions.

Khon Kaen University (KKU) serves as the apex of health care and advanced training in health sciences for northeast Thailand. As there is a diverse array of health problems specific to each region, KKU is positioned at the frontline of research on issues unique to the northeast, such as melioidosis [2] and

cholangiocarcinoma [3]. In addition, KKU also is an emerging leader in education in health and computer technology. The university offers advanced training in both areas. The Health Science Faculties comprises the Faculty of Medicine, Faculty of Dentistry, Faculty of Nursing, Faculty of Pharmaceutical Sciences, Faculty of Associated Medical Sciences, Faculty of Veterinary Medicine, and Faculty of Public Health. Training in computer technology is led by the Faculty of Engineering and Faculty of Science, with the former focusing more on hardware and the latter focusing more on software.

Building research and innovation via interdisciplinary collaboration is a key element of the mission of KKU. Despite the encouragement of the university leadership, there have not been many significant successful collaborations around health care and technology. One of the barriers was finding the opportunity to bring together individuals from the different disciplines to address key health challenges, such as improving population health literacy, tracking disease trajectory and outcomes among rural communities, and supporting the workflow of overburdened frontline providers. The prospect of developing and implementing a health care hackathon created a valuable opportunity to achieve this goal. In addition to providing this opportunity, the Harvard–Massachusetts Institute of Technology (MIT) Division of Health Science and Technology (HST) Sana team provided the experience and knowledge to conduct a hackathon, opening the door to new innovations and collaboration.

MIT Sana and Hackathons

Hackathons are intense, short, collaborative events focused on creating innovative solutions for pressing problems [4,5]. Over the last 3 years, the Laboratory for Computational Physiology (LCP) at HST, which hosts Sana, has organized more than a dozen of these events to imagine new ways technology can address critical health care challenges and develop local capacity that is fully supported by global networks.

Sana is a global consortium of academic partners with an interest in leveraging mobile health technology (mHealth) to improve health care delivery in resource-constrained settings. The hackathons have brought together students and professionals from different disciplines such as engineering, computer science, medicine, social service, public health, and business to design and develop mHealth applications.

Table 1. Health hackathons organized by the Massachusetts Institute of Technology Laboratory for Computational Physiology.

Date	Event	Location
2017		
July	Health Data Workshops	Cebu, Philippines
July	Health Datathon	Singapore
June	Mobile Health Hackathon	Mexico City, Mexico
May	Health Datathon	Sao Paulo, Brazil
April	Hacking Discrimination Hackathon	MIT[a]
March	Health Datathon	Melbourne, Australia
January	Mobile Health Hackathon	Khon Kaen, Thailand
2016		
December	Health Datathon	London, United Kingdom
October	Health Datathon	Beijing, China
September	Internet of Things Hackathon	Taipei, Taiwan
August	Hacking Mobile Health Hackathons	MIT
January	Mobile Health Hackathon	Mexico City, Mexico
2015		
October	Mobile Health Hackathon	Thessaloniki, Greece
September	Health Datathon	MIT and London, United Kingdom
July	Mobile Health Hackathon	Kampala, Uganda
June	Mobile Health Hackathon	Popayan, Colombia
2014		
September	Health Datathon	MIT, London, United Kingdom, and Paris, France
January	Health Datathon	MIT

[a]MIT: Massachusetts Institute of Technology.

Methods

Implementation of Khon Kaen University mHealth Hackathon

As an academic institution, KKU decided that the event should benefit both society and the university's internal stakeholders (university missions, faculty members, and students). With this guiding principle, the goals of the event were as follows:

- Solve public health issues in the northeast and Thailand using mHealth
- Promote collaborations between faculty members
- Provide participants with a learning opportunity that is hands-on and interdisciplinary
- Provide participants with international exposure and collaboration
- Foster research and innovation

The organizational process and structure of the event addressed these goals, along with incorporating modifications to recognize Thai culture.

Team Size and Composition

The KKU hackathon differed significantly from other hackathons in the participant recruitment and selection process. Instead of having participants joining individually, the organizing committee decided to form teams from health sciences and computer technology ahead of the event and match and merge the teams at the beginning of the event. Because assertiveness and outspokenness are not the norm for Thai people [6], this approach was aimed at reducing the stress of getting acquainted within a short amount of time and forming impromptu teams.

As the optimal team size for software development is suggested to be 7 individuals [7], the organizers decided that each team would have 2 to 3 members from health care, 2 members from computer technology, and 2 members from other related disciplines such as design and business.

Recruiting Health Science Teams

Each of the Health Science Faculties was allocated a number of teams based on its size and readiness. Each faculty was tasked with recruiting teams with the conditions that each team bring with it a specific and distinct health problem to be solved with mHealth technology and at least one health professional from the team be continually present during hack days.

Selection of teams was done internally by the leadership of each faculty, as it was felt they knew best about its disciplines and internal characteristics. The organizing committee accepted all teams that had been selected by their faculties.

Recruiting Computer Technology Teams

The Department of Computer Science and Department of Computer Engineering announced recruitment for the event in their respective departments. Initially, 35 students from computer science and 53 students from computer engineering were interested in joining the hackathon. Interested students progressed through a series of training activities that would prepare them for the event. Students who regularly attended the activities were moved to the top of the candidate list. Finally, candidates answered a short online exam and underwent an interview process. Based on student academic records, training, and the interview, the Department of Computer Science, Faculty of Science, and the Department of Computer Engineering, Faculty of Engineering, each selected 20 candidates.

The 40 candidates were then merged into 20 teams, each with one computer science student and one computer engineering student. Each pair was allowed to recruit 2 additional members from other related disciplines to complete a team of 4 members.

Matching the Teams from Health Care and Computer Technology

The organizing committee felt it was essential that the health care and engineering teams have the opportunity to find their best match, given that each project and team had specific needs and capabilities. In order to accomplish this, the organizing committee came up with an innovative approach: using a speed-dating format [8] for team introduction followed by using the US National Resident Matching Program algorithm [9] to match teams. The main goals were to allow the teams to get acquainted in a short time and match them efficiently and fairly.

To mimic the speed-dating formatting, we set up a circle of 20 tables and asked each health care team to stay inside the circle and remain stationary at its assigned table. The computer technology teams lined up outside the circles. The teams had 3 minutes to talk to each other and 30 seconds to reflect on the conversation. When the time was up, each computer technology team moved to the next table and the process began anew until every team had met all teams from the other group. Each team then ranked up to 10 teams from the other group based on its levels of preference. The rankings from all 40 teams were collected, and 2 members of the organizing committee manually applied the algorithm to match the teams. A total of 18 pairs were matched with the algorithm, and 2 pairs that could not be matched were matched with a coin toss.

Results

Bootcamp

Prior to the actual hackathon, a 2-day bootcamp was conducted. The purpose of the bootcamp was to introduce key concepts on designing mHealth innovations and methods of collaboration between clinicians and engineers, as well as address how to implement and evaluate mHealth solutions.

To provide content to a diverse audience, general topics relevant to all participants were covered in the mornings. In the afternoons, a variety of workshops was provided for participants to choose from, each addressing specific skills and/or subject matters. All lectures were given in English, with opportunities to ask questions in Thai. The language of each workshop was determined by its instructors.

Hack Days

Participants were given 48 hours of hack time to complete their projects and submit their presentation. It was not required to have a fully developed, functioning app by presentation day. Instead, the requirements were for the teams to design products

that would showcase their ideas and also show that a minimum viable product is feasible.

All teams were housed in the main hall of the KKU central library, with each having its own working station, network sockets, and power outlets. Participants brought their own electronic equipment. There was a team from KKU and mentors to help with miscellaneous requests from the participants. Refreshments were provided throughout the event.

There were both Thai and international mentors. In total, there were 4 mentors from MIT, 2 from Harvard Medical School (HMS), 1 from University of Waterloo, and 16 from KKU. Mentoring was provided throughout the hackathon and during scheduled mentor rounds. There was a board on which participants could write their problems. Mentors would periodically check the board and offer help if the posted problems were within their fields of expertise.

There were also mentor rounds and meetings throughout the event. During rounds, groups of mentors roamed the event hall to check the teams' progress and provide answers to problems. In addition, each team also had 2 scheduled mentor meetings in a private room, during which it reported its progress to a panel of mentors.

Table 2. Schedule of bootcamp and hack days.

Date	Time	Activity
January 9, 2017	9:00-9:30 AM	Introduction and opening ceremony
	9:30-10:15 AM	Solving the Problems of Health Care
	10:30-11:15 AM	Design Thinking in Global Health Informatics
	11:15 AM-12:00 PM	Bridging the Divide Between Information Technology Developers and Clinicians
	1:00-1:45 PM	Asia eHealth Information Network
	1:45-3:15 PM	Workshops in breakout rooms
January 10, 2017	9:00 -9:45 AM	The Formula for Good Health
	9:45-10:30 AM	Bringing eHealth Solutions to Market
	11:00 AM-12:00 PM	Creating a Culture of Entrepreneurship
	1:00-2:30 PM	Workshops in breakout rooms
	2:30-4:00 PM	Team introduction and matching
January 11, 2017	8:30 AM onward	Full day of hackathon
January 12, 2017		Full day of hackathon
January 13, 2017	8:30 AM	Deadline for submitting presentations
	8:40-11:20 AM	Team presentation
	11:20-11:40 AM	Deliberation of winners
	11:40 AM-12:00 PM	Award ceremonies and closing of event

Judging

There were 3 judges from the MIT Sana team, 2 judges from KKU, and 2 judges from Thai public health organizations. The main judging criteria were innovativeness, feasibility, and value of the projects. The winning team created a mobile app to optimize the logistics of moving patients around in a hospital. The first runner-up built a mobile app to incentivize blood donation and make the process more convenient for donors. The second runner-up tackled a mobile app that would allow patients on continuous peritoneal dialysis to collect data and better communicate with their caregivers.

Handling of Intellectual Properties

All participants ceded their rights over the creations to KKU under an agreement. With this agreement, KKU would be the sole entity overseeing commercialization of the products, with potential revenues to be shared between KKU and the teams. Having KKU as the sole owner of intellectual property would facilitate administrative processes regarding the projects.

One of the goals was to benefit the public as much as possible. To strike the right balance, KKU released all ideas and products from the event to the public under Creative Commons License Attribution-NonCommercial 4.0 International [10]. With this license, the public would be able to further develop the ideas and products, as long as such ventures are noncommercial.

Discussion

Building a Fruitful Collaboration

The interest, hands-on engagement, and strong support of the leadership in each of the academic departments and administration at KKU were key factors contributing to the success of the KKU mHealth hackathon. From the beginning, this level of support allowed the KKU faculty members to have the time and resources necessary for planning and organizing the event, along with building enthusiasm and participation from the entire university.

There were other important elements that supported the hackathon along with laying the foundation for ongoing collaboration between KKU and the HMS-MIT teams. These included commitment, flexibility, respect for culture and local practices, and willingness to understand each other's needs. For example, the KKU members needed to learn about the details of organizing and conducting a health care hackathon, while the HMS-MIT Sana members needed to learn about Thai health care and educational systems.

Detailed planning and careful implementation were 2 other critical elements of the successful hackathon. Use of Web-based document-sharing and Internet conference calling were critical for communication. Three KKU faculty members also had the benefit of attending an MIT Sana hackathon in Taipei, Taiwan, 4 months prior to the KKU hackathon.

Mentoring for Success

One strength of this event compared to prior hackathons was the relatively large number of mentors. With staff members from multiple departments at KKU and a large contingent of visiting mentors from the HMS-MIT Sana team, we were able to provide a diverse and large group of clinicians, engineers, and social and data scientists as mentors. There were 25 mentors in total for 20 hackathon teams.

While the diverse perspectives were certainly useful, the increasing numbers provided new challenges. We structured the mentors into small groups comprised of both Thai and international mentors. These small groups roamed the event space answering questions and reviewing the progress of the teams in order to help ensure that specific questions were answered in a timely fashion. Teams submitted questions on a request board and mentors with the appropriate expertise responded to the teams.

As the event continued, we realized that this large group of mentors required a more structured mechanism for tracking suggestions and feedback. We wanted to strike an appropriate balance between oversight and allowing teams the space to work on their own. To this end we trialed the use of a paper log at each group's table so that mentoring teams would be able to record basic notes on their interactions with a project group. This enabled the next mentoring team to identify whether issues had already been addressed and whether the team had recently received (potentially conflicting) advice from other mentors.

Unfortunately, given the impromptu addition of this system, it was not adopted across all mentoring groups. We look forward to implementing a better system of tracking the mentor/team interactions in the future and have begun brainstorming digital solutions that might be less cumbersome than paper logs.

Cultural Awareness

Unlike most hackathons where participants form teams at the beginning of the event without much of a system, we decided to be more structured with team composition. An open format could be intimidating and disorienting to inexperienced participants, and it might lead to groups without a good mix of expertise and skills. Considering that the hackathon is a novel concept for Thailand, we believe that this decision was a correct one and that it was crucial to the success of the event.

A Sense of Mission

Each team tackled a unique health issue—no 2 teams worked on the same problem. It was also required that at least 1 health care professional be present at all times during hack days to provide prompt feedback to the computer technology team. We believe that these 2 elements greatly contributed to the event, as they made the teams feel invested in their projects.

Efficient and Fair Matching

We also believe that using the speed-dating format and the US resident matching algorithm allowed for an informed, efficient, and fair matching process. Each team had a chance to learn about all teams from the other group before deciding which teams they preferred. The time it took for this process (including orientation) was only 75 minutes. And because each team was allowed to populate its own preference list before having a well-established algorithm impartially applied to it, satisfaction in the process was high.

Conclusion

Innovations in every discipline are the product of people coming together, typically with different perspectives but shared goals. Great accomplishments are seldom the work of a single individual, and yet we still focus on individual learning in schools, at the workplace, and in life. The hackathon provided a platform for the participants to teach and learn from each other. Although mentors provided the episodic one-to-many model of teaching, we observed continuous many-to-many model of learning throughout the event. We witnessed firsthand how a sense of mission and a feeling of trust within each group were critical to its success during the hackathon. None of us is as smart as all of us.

Acknowledgments

The hackathon was funded by KKU, MIT Sana, and the Center for Eye Research and Education. There was no funding for authoring this paper, and the aforementioned sponsors were not involved in approving this paper. LAC is funded by the National Institutes of Health through the R01 grant EB017205-01A1.

We would like to acknowledge the following KKU leaders and administrators who made the KKU mHealth hackathon 2017 a success: Dr Kittichai Triratanasirichai, President of KKU; Dr Nawarat Wara-aswapati Charoen, Vice President for International Affairs; Dr Denpong Soodphakdee, Vice President for Academic and Information Technology; Dr Chanchai Panthongviriyakul, Dean of the Faculty of Medicine; Dr Kanda Saikaew, Head of the Department of Computer Engineering; Dr Pipat Reungsang,

Head of the Department of Computer Science; Dr Bandit Thinkamrop; Dr Kitt Tientanopajai; Dr Wanida Kanarkard; and Dr Daris Theerakulpisut.

Conflicts of Interest

None declared.

References

1. Wibulpolprasert S, Thaiprayoon S. Thailand: good practice in expanding health coverage—lessons from the Thai health care reforms. In: Gottret P, Schieber GJ, Waters HR, editors. Good Practices in Health Financing: Lessons from Reforms in Low and Middle-Income Countries. Washington: The World Bank; 2008:355-383.
2. Limmathurotsakul D, Wongratanacheewin S, Teerawattanasook N, Wongsuvan G, Chaisuksant S, Chetchotisakd P, et al. Increasing incidence of human melioidosis in northeast Thailand. Am J Trop Med Hyg 2010 Jun;82(6):1113-1117 [FREE Full text] [doi: 10.4269/ajtmh.2010.10-0038] [Medline: 20519609]
3. Sripa B, Pairojkul C. Cholangiocarcinoma: lessons from Thailand. Curr Opin Gastroenterol 2008 May;24(3):349-356 [FREE Full text] [doi: 10.1097/MOG.0b013e3282fbf9b3] [Medline: 18408464]
4. DePasse JW, Carroll R, Ippolito A, Yost A, Santorino D, Chu Z, et al. Less noise, more hacking: how to deploy principles from MIT's hacking medicine to accelerate health care. Int J Technol Assess Health Care 2014 Jul;30(3):260-264. [doi: 10.1017/S0266462314000324] [Medline: 25096225]
5. Angelidis P, Berman L, Casas-Perez MDLL, Celi LA, Dafoulas GE, Dagan A, et al. The hackathon model to spur innovation around global mHealth. J Med Eng Technol 2016;40(7-8):392-399. [doi: 10.1080/03091902.2016.1213903] [Medline: 27538360]
6. Hofstede G. National cultures in four dimensions: a research-based theory of cultural differences among nations. Int Stud Manag Organ 1983;1(13):1-2.
7. Putnam D. Team size can be the key to a successful software project. URL: http://www.qsm.com/process_improvement_01.html [accessed 2017-06-21] [WebCite Cache ID 6rNc8HzMt]
8. Finkel EJ, Eastwick PW. Speed-dating. Curr Dir Psychol Sci 2008;17(3):193-197.
9. National Resident Matching Program. How the matching algorithm works. URL: http://www.nrmp.org/matching-algorithm/ [accessed 2017-06-21] [WebCite Cache ID 6rNeq26cL]
10. Creative Commons. Attribution–NonCommercial 4.0 International. URL: https://creativecommons.org/licenses/by-nc/4.0/legalcode [accessed 2017-06-21] [WebCite Cache ID 6rNf2e5tU]

Abbreviations

HMS: Harvard Medical School
HST: Health Sciences Technology
KKU: Khon Kaen University
LCP: Laboratory for Computational Physiology
MIT: Massachusetts Institute of Technology

Edited by G Eysenbach; submitted 22.06.17; peer-reviewed by F Kaggwa, S Van Poucke; comments to author 09.08.17; revised version received 20.08.17; accepted 20.08.17; published 16.10.17.

Please cite as:
Pathanasethpong A, Soomlek C, Morley K, Morley M, Polpinit P, Dagan A, Weis JW, Celi LA
Tackling Regional Public Health Issues Using Mobile Health Technology: Event Report of an mHealth Hackathon in Thailand
JMIR Mhealth Uhealth 2017;5(10):e155
URL: http://mhealth.jmir.org/2017/10/e155/
doi:10.2196/mhealth.8259
PMID:29038098

Original Paper

Guidelines and mHealth to Improve Quality of Hypertension and Type 2 Diabetes Care for Vulnerable Populations in Lebanon: Longitudinal Cohort Study

Shannon Doocy[1], PhD; Kenneth E Paik[2], MBA, MMSc, MD; Emily Lyles[1], MPA, MSPH; Hok Hei Tam[3], BS; Zeina Fahed[4], BS; Eric Winkler[2], BS; Kaisa Kontunen[4], MPH, MD; Abdalla Mkanna[4], MPH; Gilbert Burnham[1], MD

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Sana mHealth Group, Massachusetts Institute of Technology, Cambridge, MA, United States

[3]Sana mHealth Group, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States

[4]International Organization for Migration, Beirut, Lebanon

Corresponding Author:
Shannon Doocy, PhD
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 4105022628
Email: doocy1@jhu.edu

Abstract

Background: Given the protracted nature of the crisis in Syria, the large noncommunicable disease (NCD) caseload of Syrian refugees and host Lebanese, and the high costs of providing NCD care, the implications for Lebanon's health system are vast.

Objective: The aim of this study was to evaluate the effectiveness of treatment guidelines and a mobile health (mHealth) app on quality of care and health outcomes in primary care settings in Lebanon.

Methods: A longitudinal cohort study was implemented from January 2015 to August 2016 to evaluate the effectiveness of treatment guidelines and an mHealth app on quality of care and health outcomes for Syrian and Lebanese patients in Lebanese primary health care (PHC) facilities.

Results: Compared with baseline record extraction, recording of blood pressure (BP) readings (−11.4%, P<.001) and blood sugar measurements (−6.9%, P=.03) significantly decreased following the implementation of treatment guidelines. Recording of BP readings also decreased after the mHealth phase as compared with baseline (−8.4%, P=.001); however, recording of body mass index (BMI) reporting increased at the end of the mHealth phase from baseline (8.1%, P<.001) and the guidelines phase (7.7%, P<.001). There were a great proportion of patients for whom blood sugar, BP, weight, height, and BMI were recorded using the tablet compared with in paper records; however, only differences in BMI were statistically significant (31.6% higher in app data as compared with paper records; P<.001). Data extracted from the mHealth app showed that a higher proportion of providers offered lifestyle counseling compared with the counseling reported in patients' paper records (health diet counseling; 77.3% in app data vs 8.8% in paper records, P<.001 and physical activity counseling and 59.7% in app vs 7.1% in paper records, P<.001). There were statistically significant increases in all four measures of patient-provider interaction across study phases. Provider inquiry of medical history increased by 16.6% from baseline following guideline implementation and by 28.2% from baseline to mHealth implementation (P<.001). From baseline, patient report of provider inquiry regarding medication complications increased in the guidelines and mHealth phases by 12.9% and 59.6%, respectively, (P<.001). The proportion of patients reporting that providers asked other questions relevant to their illness increased from baseline through guidelines implementation by 27.8% and to mHealth implementation by 66.3% (P<.001). Follow-up scheduling increased from baseline to the guidelines phase by 20.6% and the mHealth phase by 39.8% (P<.001).

Conclusions: Results from this study of an mHealth app in 10 PHC facilities in Lebanon indicate that the app has potential to improve adherence to guidelines and quality of care. Further studies are necessary to determine the effects of patient-controlled health record apps on provider adherence to treatment guidelines, as well as patients' long-term medication and treatment adherence and disease control.

(JMIR Mhealth Uhealth 2017;5(10):e158) doi:10.2196/mhealth.7745

KEYWORDS

mHealth; hypertension; diabetes mellitus; chronic disease; Lebanon, Syria; refugees

Introduction

An estimated 4.8 million Syrians have fled the conflict to neighboring countries and are registered or awaiting registration with the United Nations High Commissioner for Refugees (UNHCR), in addition to a population of unregistered refugees unknown in number [1]. As of January 2017, over one million Syrian refugees were registered with UNHCR in Lebanon [1]. With an estimated 183 refugees per 1000 inhabitants at the end of 2015, Lebanon hosts the highest ratio of refugees-to-host population worldwide [2]. The humanitarian response in Lebanon is coordinated through an interagency mechanism established by UNHCR and the Lebanese government, integrating refugee assistance into existing clinics. Delivery of health services for Syrian refugees is based on a primary health care (PHC) strategy. Syrian refugees can utilize primary health care services paying subsidized rates at designated existing primary health care centers and primary level facilities across Lebanon, unless they choose to seek care at private clinics [3,4]. Delivery of noncommunicable disease (NCD) treatment for Syrian refugees and vulnerable Lebanese not seeking care in the private sector is based on routine care in primary health facilities with referral to secondary and tertiary care for specialist management.

Both Lebanese and Syrian populations are in the late stages of the epidemiologic transition from communicable, maternal, neonatal, and nutritional conditions to NCDs. In Lebanon, both the host community and refugee populations suffer from high NCD burdens [5,6]. Type 2 diabetes prevalence has been estimated at 7.4% in Syria and 14.4% in Lebanon [7]. Previous reports have estimated regional prevalence of hypertension at 29.5% in Syria and for Lebanon variously at 24.9% and 28.8% [8-10]. Ischemic heart disease and stroke, for which hypertension and diabetes have substantially increased risk, are the leading causes of death in Lebanon and aside from conflict-related death, in Syria as well [11,12]. Moreover, based on the 2012 age-specific mortality risks throughout their lifetime, the probability of an individual aged between 30 and 70 years dying from cancer, cardiovascular disease, chronic respiratory disease, or diabetes is 12% in Lebanon and 19% in Syria; figures matched only by risks of conflict-related death in Syria [11,12]. Management of NCDs can be difficult and requires continuity of care, which is difficult for refugees and poses challenges to health services and systems. The burden placed on Lebanon's highly fragmented and privatized health system by refugee influx is immense though not unique in the new global displacement environment [13]. Increasingly, displaced populations are urban and from low- and middle-income countries where NCDs constitute a significant burden of disease. Not only the numbers but also the complexity of conditions pose challenges to health systems addressing the needs of both refugee and hosts with NCDs. The practice pattern in which persons with even mild hypertension are seen by cardiologists and persons with well-controlled mild diabetes consult endocrinologists rather than primary care physicians increases the complexity and costs of care. Limited resource availability has prioritized care to PHC conditions, limiting more expensive specialist care [3,14,15]. We undertook a study to evaluate the effectiveness of treatment guidelines and an mHealth app on quality of care and health outcomes in primary care settings.

Methods

Study Design

A longitudinal cohort study was implemented from January 2015 to August 2016 in primary health facilities in Lebanon that serve both Syrian refugees and Lebanese. Its two research aims were (1) to develop, adapt, and test existing standards and guidelines for treatment, including counseling, of persons with hypertension and type 2 diabetes (or both) and (2) to evaluate the effectiveness of an mHealth tool. Standard best-practice guidelines were adapted to the local context using national protocols, prescribing practices, and the primary care context where they would be applied. [16-18]. Providers were subsequently trained on guidelines and provided with written materials to support clinical decision making. The mHealth app included a personally controlled health record (PCHR), informational printouts for patients on prescriptions, and lifestyle behaviors and served as an electronic medical record and decision support tool for providers. If patients move locations without their medical records, key diagnostic and treatment elements are available from the patient's cell phone subscriber identity module card, which constitutes the PCHR. The mHealth tool has the potential to improve quality and continuity of care, health literacy, mobility of medical records, and health outcomes for patients. Providers were trained in use of the app, and support was provided to health facilities for its implementation [19]. The study used a phased introduction of the two interventions over 20 months with longitudinal measurement of outcomes.

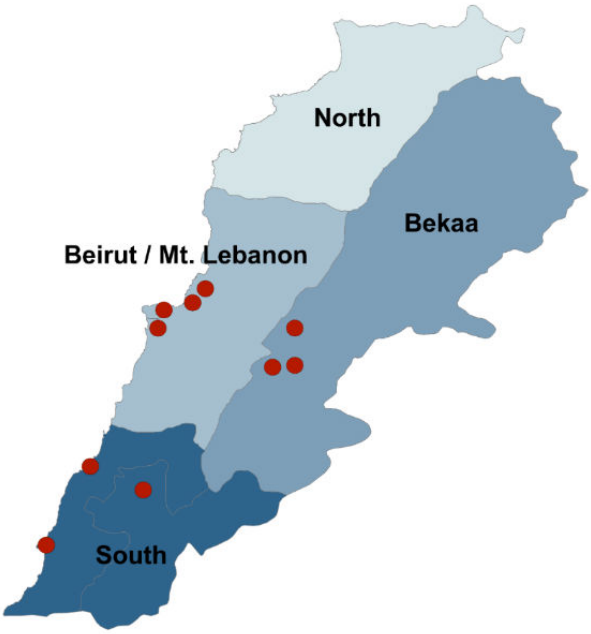
Study Participants

Participants consisted of patients at 10 health care centers in Lebanon supported by the International Organization for Migration or the International Medical Corps in the South (n=3), Bekaa (n=3), and Beirut and Mount Lebanon (n=4) governorates (Figure 1). Patients at these locations were predominantly Lebanese and Syrian refugees. Individuals without a diagnosis of hypertension or type 2 diabetes, those aged less than 40 years, and adults lacking capacity to independently participate in interviews were excluded.

A total of 1020 participants were enrolled and 793 (77.75%) completed the study. Sample size calculations were based on the estimated proportion of providers adhering to treatment guidelines, with an assumed baseline rate of 50% for adherence to guidelines (the most conservative rate that would ensure the ability to detect significant differences from all other rates).

This is a reasonable assumption given that proposed guidelines did not differ substantially from other best practice guidelines; thus, patients being enrolled at baseline could already be on recommended treatment. Sample size calculations were performed using Stata 13 (StataCorp LLC), assumed alpha=.05 and beta=.20 (power=0.80), and were one-sided based on the assumption that quality of care will not decrease because of the intervention. The final sample of 793 participants was sufficient to detect increases ≥5.0% for provider adherence to guidelines.

Figure 1. Participating primary health centers.



Study Procedures and Outcome Measures

This study was designed using a mixed-methods approach with qualitative and quantitative data collected throughout. Patients were recruited at clinics, and if they indicated willingness to participate, a follow-up phone call was made. This verified consent, and a baseline interview collected information on demographic characteristics; medical history and recent care-seeking behaviors; and knowledge, attitudes, and practices related to type 2 diabetes and hypertension. Following enrollment, medical record reviews were also conducted for each patient, recording information related to provider compliance with guidelines and quality of care. Additional information was collected on frequency of clinic visits, patient status (death and loss to follow-up), and disease-specific patient outcomes (complications and adverse events of hypertension and type 2 diabetes). Data from phone interviews and record reviews were collected at the end of each study phase (guidelines and mHealth). In addition, a subset of patients visiting study facilities during the course of the study were telephoned within 10 days of their visit to complete a brief exit interview.

Clinical Measurements

Clinical measurements including height, weight, blood pressure (BP), glycated hemoglobin (HbA1c), fasting blood sugar, and random blood sugar were extracted from patient records at baseline, following implementation of the treatment guidelines, and after implementation of the PCHR. At the end of the mHealth intervention study phase, clinical measurements were also extracted from the PCHR database to triangulate facilities' record keeping with data entered in the PCHR by providers.

Patient-Provider Interaction

The quality of patient-provider clinical interactions was assessed based on patient reports from exit interviews conducted during each study phase with a subset of patients that visited a study facility. As with clinical measurements, data from the PCHR was used to compare patient report of clinical interactions with that reported by providers in the app. Interactions were evaluated based on four key indicators of providers' compliance with treatment guidelines: (1) provider inquiry of medical history, (2) query about complications with prescribed medication, (3) prompting for questions from the patient, and (4) recommending follow-up or referral care. Additionally, clinical interactions were evaluated based on report of lifestyle counseling on smoking, alcohol consumption, physical activity, and dietary patterns.

Medication Prescription and Use

Medication prescription and use were assessed and compared using data obtained both through patient self-report during phone interviews conducted in each study phase, as well as documentation in patients' health facility records.

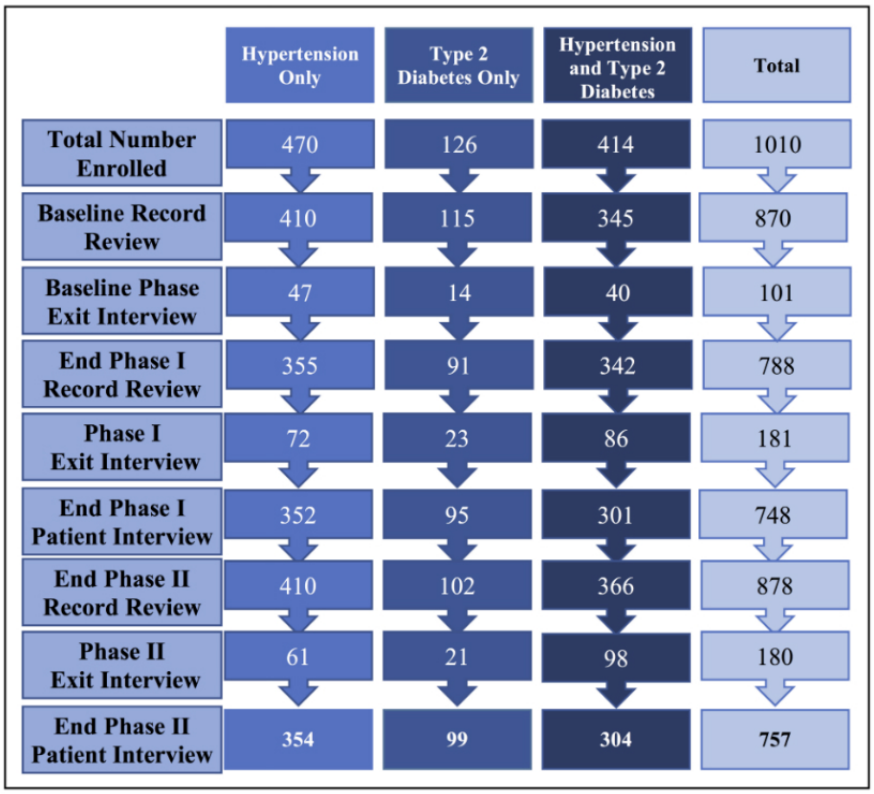
Analysis

Data were collected with tablets using the Magpi mobile data platform by DataDyne LLC (Washington, DC) and analyzed using Stata 13 (College Station, TX) using descriptive statistics and standard methods for comparison of means and proportions. BP readings monitored control among hypertensive patients, and the HbA1c test was the preferred measure for classifying type 2 diabetic patients; when not available, random or fasting

blood sugar was used [20,21]. A sequenced process-based classification used patient records, clinical data, and prescriptions to assign a uniform diagnosis category to patients in cases where reporting was inconsistent over time. A total of 8 patients remained with an unclassified diagnosis and were subsequently dropped from final analysis to ensure reliable reporting by condition. Utilization of the mHealth app by practitioners was low. A total of 154 records were extracted from the app dataset, whereas a total of 878 record reviews and 761 patient interviews were completed in the mHealth phase (Figure 2). Differences in patient characteristics and condition control status were examined using chi-square and *t* test methods. An immediate form of two-sample tests of proportions was performed using the Stata *prtesti* command to determine whether the proportions in the mHealth app and paper records were statistically different.

This study was approved by the ministry of public health in Lebanon and the institutional review board at the Johns Hopkins Bloomberg School of Public Health.

Figure 2. Patient follow-up and response rates.



Results

Clinical Measurements

Clinical measurements extracted from patient records and hypertension and type 2 diabetes control data are presented in Table 1. Compared with baseline data, significant declines in reporting of BP (−11.4%, *P*<.001) and blood sugar (−6.9%, *P*=.03) measurements were observed following implementation of treatment guidelines. Recording of systolic and diastolic BP measurements also declined after the mHealth phase as compared with baseline (−8.4%, *P*=.001); however, body mass index (BMI) reporting increased at the end of the mHealth phase from both baseline (8.1%, *P*<.001) and the end of the guidelines phase (7.7%, *P*<.001). Baseline clinical test results included all information in clinic records, regardless of when it was reported; because reporting is not time bound, changes in completeness of reporting are difficult to interpret because values could have been reported at one of a number of prior visits. Changes in clinical measurements and the control of hypertension and type 2 diabetes were not significant in the mHealth phase, likely because of short implementation time and challenges with provider uptake of the app.

Table 1. Patient biometric health measures from the noncommunicable disease (NCD) guidelines and mobile health (mHealth) records for refugees in the Lebanon study.

Parameter	Baseline		Phase I[a]		Phase II[b]		Change by phase		
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	Phase I versus baseline	Phase II versus baseline	Phase II versus phase I
							P value	*P* value	*P* value
Body mass index (BMI)	**N=870**		**N=789**		**N=878**				
Total patients with BMI measured	67 (7.7)	6.0-9.7	64 (8.1)	6.3-10.2	139 (15.8)	13.5-18.4	.76	<.001	<.001
	N=67		**N=64**		**N=139**				
Median	32.8		33		31.5				
Mean	33.5	31.9-35.1	34	32.3-35.8	32.1	31.1-33.1	.65	.13	.04
BMI (normal)[c]	5 (8)	2.5-16.6	5 (8)	2.6-17.3	12 (8.6)	4.5-14.6	.94	.78	.85
BMI (overweight)[d]	15 (22)	13.1-34.2	14 (22)	12.5-34.0	42 (30.2)	22.7-38.6	.94	.24	.22
BMI (obese)[e]	47 (70)	57.7-80.7	45 (70)	57.6-81.1	85 (61.2)	52.5-69.3	.98	.21	.21
Hypertension	**N=755**		**N=697**		**N=776**				
Total hypertension patients with blood pressure measured	371 (49.1)	45.5-52.8	263 (37.7)	34.1-41.4	316 (40.7)	37.2-44.3	<.001	.001	.24
Blood pressure	**N=371**		**N=263**		**N=316**				
Controlled blood pressure (BP)[f]	238 (64.2)	59.0-69.0	183 (69.6)	63.6-75.1	223 (70.6)	65.2-75.5	.15	.08	.80
Uncontrolled systolic BP[g]	81 (21.8)	17.7-26.4	42 (16.0)	11.8-21.0	59 (18.7)	14.5-23.4	.07	.31	.40
Uncontrolled diastolic BP[h]	7 (1.9)	0.8-3.8	5 (1.9)	0.6-4.4	6 (1.9)	0.7-4.1	.99	.99	.99
Uncontrolled BP[i]	45 (12.1)	9.0-15.9	33 (12.5)	8.8-17.2	28 (8.9)	6.0-12.6	.88	.17	.15
Diabetes	**N=460**		**N=433**		**N=468**				
Diabetes patients with blood test results[j]	173 (37.6)	33.2-42.2	133 (30.7)	26.4-35.3	159 34.0)	29.7-38.5	.03	.25	.30
Diabetes control[k]	**N=173**		**N=133**		**N=159**				
Controlled	78 (45.1)	37.5-52.8	56 (42.1)	33.6-51.0	83 (52.2)	44.1-60.2	.60	.20	.09
Uncontrolled	95 (54.9)	47.2-62.5	77 (57.9)	49.0-66.4	76 (47.8)	39.8-55.9			

[a]Guideline implementation.

[b]mHealth implementation.

[c]BMI<25 kg/m^2 normal.

[d]BMI>25kg/m^2 overweight.

[e]BMI>30kg/m^2 obese.

[f]Controlled: BP<140/90.

[g]Uncontrolled: Systolic BP>140 (Diastolic BP<90).

[h]Uncontrolled: Diastolic BP>90 (Systolic BP<140).

[i]Uncontrolled: BP>140/90.

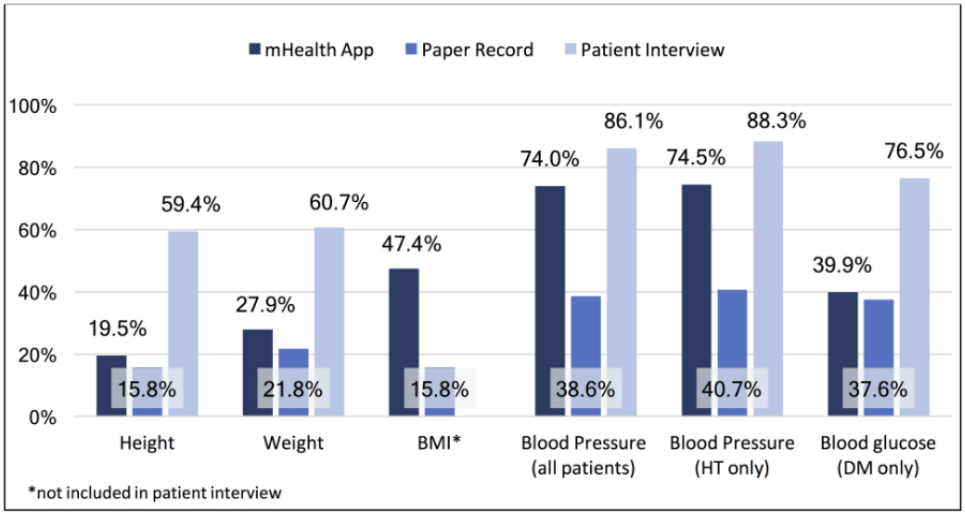
[j]Includes HbA1c, FBS, RBS, or any combination of those tests.

[k]Based on results from either HbA1c, FBS, or RBS; if multiple tests available preference is given first to HbA1c (controlled defined as <7.0%), then FBS (controlled defined as <120mg/dL), then RBS (controlled defined as <100 mg/dL).

Comparison of data reported in the mHealth app with paper records and patient interviews during app implementation is presented in Figure 3. Comparing information reported in paper records following implementation of the mHealth app with data extracted directly from the app, BP measures were reported for a substantially larger proportion of patients in the app (114/154, 74.0% patients with app data vs 339/878, 38.6% patients with paper records, *P*<.001). Similarly, reporting of weight, height,

and BMI were all more frequently reported with the app than with paper records as follows: weight, 43/154 (28%) patients from app data versus 191/878 (21.8%) from paper records, *P*=.10; height, 30/154 (19%) patients from the app versus 139/878 (15.8%) from paper records, *P*=.25; BMI, 73/154 (47%) patients from the app versus 139/878 (15.8%) from paper records, *P*<.001. Among hypertensives, BP readings were more commonly reported with the app than with paper records (114/153, 75% patients from the app vs 776/878, 40.7% patients from paper records; *P*=.24). Among type 2 diabetics, blood sugar tests were reported for a slightly larger proportion of patients with the app than with patient records (61/153, 39.9% of patients from app data vs 159/468, 34.0% from paper records; *P*=.19). Higher reporting by clinicians using the app supports the likelihood that mixed results regarding changes in provider adherence to guidelines measured following the mHealth phase are because of poor reporting with paper records more than poor performance of the app. Furthermore, over twice as many patients reported that measurement of weight, height, BP, and blood glucose had been taken than the mHealth app and/or paper records showed, suggesting that care quality may be better in actuality than as reflected by completeness of reporting measures.

Figure 3. Clinical iIndicator measurement by reporting source.



Patient-Provider Interaction

Statistically significant increases were detected in all four measures of patient-provider clinical interactions (Table 2). The proportion of patients reporting that the provider took a medical history during the enrollment phase (72/101 patients, 71.3%) increased by 16.6% to 87.9% (160/182) patients in the guideline phase and by 28.2% from enrollment to 99.4% (179/180) patients in the mHealth phase (*P*<.001). Just over a third (36/100, 36%) of patients reported that the provider asked about medication complications at the most recent care visit during the enrollment phase. In the guidelines and mHealth phases, this increased from enrollment by 12.9% to 48.9% (89/182) patients and by 59.6% to 95.6% (172/180) patients, respectively (change from enrollment to guidelines phase *P*=.04; change from enrollment to mHealth phase, *P*<.001). The proportion of patients reporting that providers asked other questions relevant to their illness increased from 32.0% (32/100) patients during the enrollment phase to 59.8% (107/179) patients during the guidelines phase and 98.3% (177/180) patients in the mHealth phase (respective increases of 27.8% and 66.3%, *P*<.001). A significantly higher proportion of patients also reported providers scheduling a follow-up appointment or being referred for specialty care, from 58.0% (58/100) patients in the enrollment phase to 78.6% (143/182) patients in the guidelines phase and 97.8% (176/180) patients in the mHealth phase (respective increases of 20.6% and 39.8%, *P*<.001).

Patient report of provider counseling about lifestyle behaviors such as smoking, alcohol consumption, physical activity, and dietary patterns also improved (Table 2). However, the provider's reports of counseling carried out significantly differed in patient records and the mHealth app. Data extracted from the mHealth app showed a much higher proportion of providers offering lifestyle counseling as compared with notations in patient records. Smoking cessation counseling was reported for 16.9% (26/154) patients from the app data versus 11.4% (96/844) patients from paper records (*P*=.06). Much larger differences were observed in health dietary habit counseling (119/154, 77.3% patients from app data vs 77/878, 8.8% from paper records; *P*<.001) and physical activity counseling (92/154, 59.7% patients from app data vs 62/878, 7.1% from paper records; *P*<.001).

Table 2. Quality of interaction with providers reported by patients in the noncommunicable disease (NCD) guidelines and mobile health (mHealth) records for refugees in Lebanon study. Data reported by patients in exit interviews were conducted via phone.

Parameter	Baseline (N=101)		Phase I[a] (N=181)		Phase II[b] (N=180)		Change comparison		
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	Phase I versus baseline	Phase II versus baseline	Phase II versus phase I
							P value	*P* value	*P* value
Provider interaction									
Asked about medical history	72 (71.3)	61.4-79.9	160 (87.9)	82.3-92.3	179 (99.4)	96.9-100	<.001	<.001	<.001
Asked about complications with medications	36 (36.0)	26.6-46.2	89 (48.9)	41.4-56.4	172 (95.6)	91.4-98.1	.04	<.001	<.001
Asked other questions	32 (32.0)	23.0-42.1	107 (59.8)	52.2-67.0	177 (98.3)	95.2-99.7	<.001	<.001	<.001
Provided follow-up appointment or referral	58 (58.0)	47.7-67.8	143 (78.6)	71.9-84.3	176 (97.8)	94.4-99.4	<.001	<.001	<.001
Lifestyle counseling received									
Quit or stop using tobacco	32 (31.7)	22.8-41.7	79 (44.1)	36.7-51.7	157 (87.2)	81.4-91.7	.04	<.001	<.001
Reduce salt consumption	56 (55.4)	45.2-65.3	147 (82.1)	75.7-87.4	172 (95.6)	91.4-98.1	<.001	<.001	<.001
Fruit and vegetable consumption	48 (47.5)	37.5-57.7	141 (78.8)	72.0-84.5	172 (95.6)	91.4-98.1	<.001	<.001	<.001
Reduce fat consumption	56 (55.4)	45.2-65.3	150 (83.8)	77.6-88.9	172 (95.6)	91.4-98.1	<.001	<.001	<.001
Engage in physical activity	43 (43.0)	33.1-53.3	138 (77.1)	70.2-83.0	167 (92.8)	88.0-96.1	<.001	<.001	<.001
Lose weight	31 (31.0)	22.1-41.0	120 (67.0)	59.6-73.9	155 (86.1)	80.2-90.8	<.001	<.001	<.001

[a]Guideline implementation.

[b]mHealth implementation.

Medication Prescription and Use

Medication compliance and other compliance variables were also used to evaluate performance and outcomes related to guideline training and app adoption (Table 3). The proportion of patients reporting receiving prescriptions of medication for hypertension and type 2 diabetes was consistently high, exceeding 90% at baseline and in both study phases (Figure 4). On the basis of reporting in patient records, there was a small but significant increase in patients prescribed medication for hypertension from baseline to the end of the guidelines phase (6.6% increase, *P*=.003) and from baseline to the end of the mHealth phase (5.1% increase, *P*=.02). Unlike notations in patient records, the proportion of patients self-reporting being prescribed hypertension medication decreased significantly from baseline to the end of the guidelines phase (9.8% decrease, *P*<.001); however, this proportion significantly increased in the mHealth phase by 8.9% (*P*<.001). The proportion of patients self-reporting current use of hypertensive medications decreased significantly from baseline to the end of the guidelines phase (3.9% decrease, *P*<.001) and from baseline to the end of the mHealth phase (2.3% decrease, *P*=.02).

Among patients with type 2 diabetes, there was a significant increase in medication prescription in patient records from baseline to the end of the guidelines phase (5.6% increase, *P*=.047) and from baseline to the end of the mHealth phase (10.1% increase, *P*<.001). Unlike in patient records, the proportion of patients self-reporting being prescribed medication decreased significantly from baseline to the end of the guidelines phase (3.9% decrease, *P*<.001); however, this proportion significantly increased in the mHealth phase by 2.9% (*P*=.03). The proportion of type 2 diabetics reporting current diabetes medication use decreased significantly from baseline to the end of the guidelines phase (6.3% decrease, *P*<.001) and from baseline to the end of the mHealth phase (3.8% decrease, *P*=.02).

Overall, medication compliance was good among both hypertensive and type 2 diabetic patients across follow-up. The proportion of hypertensive patients reporting they had stopping prescribed medication for 2 weeks or longer in the 3 months preceding interview was highest at baseline (70/755, 9.3%) and lowest at the end of the guidelines phase (49/604, 8.1%). Interruptions in diabetes medication was highest at baseline and at the end of the guidelines phase (38/506, 7.5% and 29/383, 7.6%, respectively) and lowest at the end of the mHealth phase (14/256, 5.5%). Observed changes in interruption of medication for hypertension or diabetes were not significant among any of the study phases. Reasons for stopping medication were similar by condition and across the study periods. Cost was the primary reason for stopping medication (62.9%-74.4%, depending on the condition and study phase). The other common reasons were advice from the provider and the patient perception that that their condition had improved, which is particularly challenging as, given that hypertension and diabetes have few if any symptoms, patients likely ascribe symptoms to their disease that are not related.

Table 3. Medication and compliance among patients in the noncommunicable disease (NCD) guidelines and mobile health (mHealth) records for refugees in Lebanon study.

Parameter	Baseline		Phase I[a]		Phase II[b]		Change comparison		
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	Phase I versus baseline	Phase II versus baseline	Phase II versus phase I
							P value	*P* value	*P* value
Hypertension medication	**N=755**		**N=697**		**N=776**				
All hypertension patients prescribed medication for hypertension[c]	550 (72.8)	69.5-76.0	554 (79.5)	76.3-82.4	605 (78.0)	74.9-80.8	.003	.02	.48
Not reported in patient record[c]	43 (5.7)	4.2-7.6	92 (13.2)	10.8-15.9	35 (4.5)	3.2-6.2	<.001	.29	<.001
Self-reported hypertension medication	**N=873**		**N=652**		**N=418**				
Ever prescribed medication	873 (100)	99.5-100	588 (90.2)	87.6-92.4	414 (99.0)	97.6-99.7	<.001	.004	<.001
	N=754		**N=605**		**N=413**				
Currently taking hypertension medication[d]	740 (98.1)	96.9-99.0	570 (94.2)	92.0-95.9	396 (95.9)	93.5-97.6	<.001	.02	.24
Stopped taking medicines for 2+ weeks in the past 3 months[d]	70 (9.3)	7.3-11.6	49 (8.1)	6.1-10.6	35 (8.6)	6.0-11.7	.45	.69	.80
Noncompliance[e]	**N=76**		**N=50**		**N=36**				
When medication was stopped									
Stopped taking in Syria	9 (11.8)	5.6-21.3	10 (20.0)	10.0-33.7	5 (13.9)	4.7-29.5	.21	.76	.47
Taking in Syria, stopped in Lebanon	36 (47.4)	35.8-59.2	13 (26.0)	14.6-40.3	15 (41.7)	25.5-59.2	.02	.58	.13
Started taking in Lebanon but stopped	31 (40.8)	29.6-52.7	27 (54.0)	39.3-68.2	16 (44.4)	27.9-61.9	.15	.72	.39
Diabetes medication	**N=460**		**N=433**		**N=468**				
% of all diabetes patients prescribed medication for diabetes[c]	343 (74.6)	70.3-78.5	347 (80.1)	76.1-83.8	396 (84.6)	81.0-87.8	.047	<.001	.08
Not reported in patient record[c]	28 (6.1)	4.1-8.7	54 (12.5)	9.5-16.0	16 (3.4)	2.0-5.5	.001	.06	<.001
Self-reported diabetes medication	**N=537**		**N=394**		**N=260**				
Ever prescribed medication	536 (99.8)	98.9-100	378 (95.9)	93.5-97.7	257 (98.8)	96.7-99.8	<.001	.07	.03
	N=506		**N=384**		**N=259**				
Currently taking diabetes medication[d]	488 (96.4)	94.4-97.9	346 (90.1)	86.7-92.9	240 (92.7)	88.8-95.5	<.001	.02	.26
Stopped taking medicines for 2+ weeks in the past 3 months[d]	38 (7.5)	5.4-10.2	29 (7.6)	5.1-10.7	14 (5.5)	3.0-9.0	.97	.29	.30
Noncompliance[e]	**N=43**		**N=35**		**N=15**				
When medication was stopped									
Stopped taking in Syria	6 (14.0)	5.3-27.9	7 (20.0)	8.4-36.9	5 (33.3)	11.8-61.6	.48	.10	.32

Parameter	Baseline		Phase I[a]		Phase II[b]		Change comparison		
	n (%)	95% CI	n (%)	95% CI	n (%)	95% CI	Phase I versus baseline	Phase II versus baseline	Phase II versus phase I
							P value	*P* value	*P* value
Taking in Syria, stopped in Lebanon	22 (51.2)	35.5-66.7	9 (25.7)	12.5-43.3	3 (20.0)	4.3-48.1	.02	.04	.67
Started taking in Lebanon but stopped	15 (34.9)	21.0-50.9	19 (54.3)	36.6-71.2	7 (46.7)	21.3-73.4	.09	.43	.63

[a]Guideline implementation.

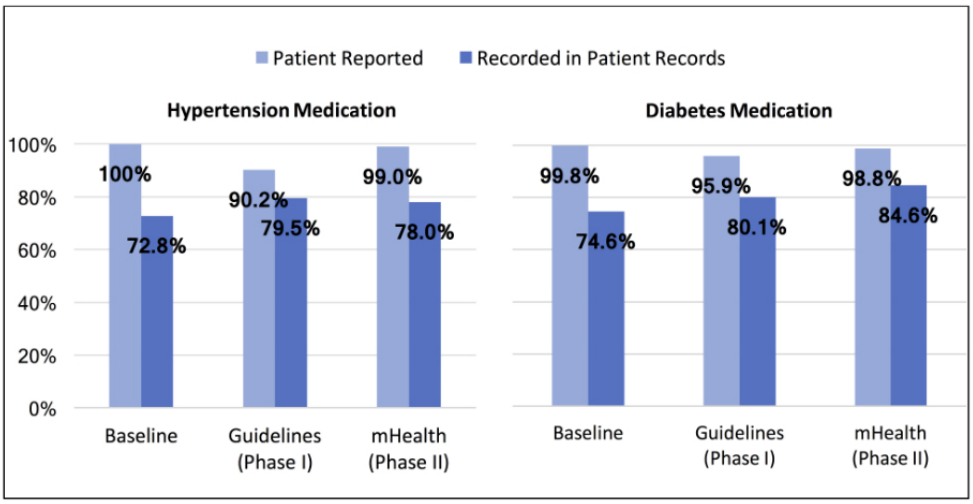
[b]mHealth implementation.

[c]As reported in patient health records.

[d]Among those prescribed medication.

[e]Among only patients that stopped taking medication in the past 3 months.

Figure 4. Medication prescription by reporting method.



Discussion

Principal Findings

Consistent and complete reporting is essential to monitor changes and trends in clinical measurements for diabetic and hypertensive patients. Notwithstanding relatively low provider uptake of the app, reporting of nearly all clinical measures was improved when the provider used the app rather than written patient records. Data extracted from the mHealth app also showed a greater proportion of providers offering lifestyle counseling as compared with counseling reported in paper medical records kept by health facilities, and statistically significant improvements were observed in all four measures of patient-provider interaction. This clearly demonstrated the advantages of electronic reporting and the potential for mHealth apps to improve quality of clinical care for chronic NCDs. Despite the possible benefits of mHealth for improving case management of hypertension and type 2 diabetes, there were difficulties in developing and deploying new technologies that diminished the utilization and potential benefits of the mHealth app.

The challenges in this study are neither unique to the project nor to the context. Data quality in electronic health records (EHRs), particularly completeness of reporting, is indispensable for the associated decision support components to prove effective. Poor reporting observed in this project were similarly observed in a recent trial incorporating the Screening Tool of Older People's Prescriptions prescribing criteria in a primary care EHR, which demonstrated the need for continued assessment of data quality and improvement on potentially inappropriate prescription rates in community primary care settings [22]. A 2013 study of implementation of national guidelines and an associated structured type 2 diabetes and hypertension patient record reported poor use of structured record by providers as a primary explanation for null benefit of the intervention [23]. Other barriers previously documented include cost, language, literacy, availability or connectivity issues, and perceived increase in workload; connectivity, language, and workload increase presented challenges in this study [24,25]. The aim of this project focused on both PCHR app development and pilot evaluation of the app. However, allowing a longer time period to develop and test the app, followed by a subsequent pilot test, would have been a more appropriate design if time permitted. As such, mixed findings on the results of the mHealth app use should not minimize the consideration of the app's potential effectiveness.

The nature of mHealth interventions vary widely, ranging from health promotion and disease surveillance to remote monitoring, care support, and decision support tools [26]. Previous research has primarily focused on remote monitoring and care support tools, and while there is a considerable body of literature on the design and implementation of personal health records, considerably less evidence is available for decision support functions, particularly in low- and middle-income countries [26,27]. EHRs have previously shown effectiveness in improving type 2 diabetic patient health outcomes and clinical practice in developed countries but do not adequately capture the potential added benefits of provider decision support elements as were incorporated in the PCHR app developed for this study [28]. The PCHR app developed for this study might lead to better uptake and prove more effective in other settings, with providers more open to newer technology, with fewer reporting requirements, without electronic information management systems, and where providers were more open to changing their clinical practice behaviors. The patient-controlled portability component may also improve patient knowledge of their condition and continuity of care, in particular in the context of migration, two outcomes which were not assessed in this study.

Limitations

Comparison of completeness of reporting across study phases may have underestimated changes in patient and provider practices in the guidelines and mHealth phases where all available information in the patient record was included at baseline, regardless of what was recorded at the most recent visit. Simultaneous development and introduction of the app led to frustration among users when the app did not perform as expected, requiring frequent software updates, which reduced provider enthusiasm. Another barrier to uptake was multiple reporting requirements and electronic record systems, which led to the perception that the app was redundant (despite dissimilarities to existing systems in most cases). Finally, including patients and providers from only 10 health facilities limits representativeness of findings, and the results may not be generalizable to elsewhere in Lebanon or other settings.

Conclusions

The mHealth app was successful in improving some quality of care indicators, indicating there is potential for clinical decision making support tools to enhance capacity for NCD care in PHC centers. Recording rates of BMI improved with use of the mHealth app; however, there was a decline in the recording of BP and blood sugar levels; recording rates for all three measures were higher in the mHealth app than in paper records. Patient-provider interactions, life style counseling, and scheduling of follow-up appointments improved with use of the mHealth app, suggesting there were some improvements in quality of care. Only small improvements in the proportion of patients with controlled hypertension and diabetes were observed between baseline and the end of the mHealth phase, and these differences were not statistically significant.

Results from this study of an mHealth app in 10 PHC facilities in Lebanon indicate the app has potential to improve adherence to guidelines and quality of care. Greater support to service providers during the adoption of the apps, customization of the apps for specific settings, and longer follow-up periods may aid in better characterizing possible benefits of this and other mHealth apps for NCD management. Further studies are necessary to determine the effects of this and similar PCHR apps on provider adherence to treatment guidelines, as well as patients' long-term medication and treatment adherence and disease control. Additional testing in less developed settings, including both rural locations and emergency contexts, will help provide evidence on the potential of these apps and factors associated with uptake and effectiveness across a broader range of contexts. Expanding the evidence on mHealth apps so that it is sufficient to inform decision making on adoption of these tools is essential, given their potential benefits.

Acknowledgments

This research was funded by Research for Health in Humanitarian Crisis. The funding body had no role in the design or implementation of the research and did not participate in analysis or presentation of findings. The authors would like to express their utmost gratitude to the staff at the 10 health facilities that participated in this study—without their dedication and support, this work would have not been possible. They are also grateful to the data collection team and their many colleagues at the Johns Hopkins School of Public Health, the International Organization for Migration, and the Massachusetts Institute of Technology Sana mHealth Group for their contributions to the project.

Conflicts of Interest

None declared.

References

1. United Nations High Commissioner for Refugees. Data.unhcr. Syria regional refugee response URL: http://data.unhcr.org/syrianrefugees/regional.php [WebCite Cache ID 6p9yjYPqk]
2. United Nations High Commissioner for Refugees. Unhcr. 2016. Global trends forced displacement in 2015 URL: http://www.unhcr.org/576408cd7.pdf [accessed 2017-03-22] [WebCite Cache ID 6pA38AO2W]
3. World Bank. 2016. Lebanon-emergency primary healthcare restoration project URL: http://documents.worldbank.org/curated/en/185271468266958778/pdf/PAD12050PAD0P15264600PUBLIC00Box391428B.pdf [WebCite Cache ID 6pA06nNYg]

4. Ministry of Public Health of Lebanon. Moph. 2015. Health response strategy: a new approach in 2015 and beyond URL: http://www.moph.gov.lb/userfiles/files/Strategic%20Plans/HRS-DRAFT8.pdf [accessed 2017-03-22] [WebCite Cache ID 6pA0GdlC3]
5. Amara AH, Aljunid SM. Noncommunicable diseases among urban refugees and asylum-seekers in developing countries: a neglected health care need. Glob Health 2014 Apr 03;10:24 [FREE Full text] [doi: 10.1186/1744-8603-10-24] [Medline: 24708876]
6. Shahraz S, Forouzanfar MH, Sepanlou SG, Dicker D, Naghavi P, Pourmalek F, et al. Population health and burden of disease profile of Iran among 20 countries in the region: from Afghanistan to Qatar and Lebanon. Arch Iran Med 2014 May;17(5):336-342 [FREE Full text] [Medline: 24784862]
7. IDF. 2014. IDF diabetes atlas 2014 URL: http://www.idf.org/sites/default/files/DA-regional-factsheets-2014_FINAL.pdf [accessed 2017-03-22] [WebCite Cache ID 6pA0JG8xT]
8. Tailakh A, Evangelista LS, Mentes JC, Pike NA, Phillips LR, Morisky DE. Hypertension prevalence, awareness, and control in Arab countries: a systematic review. Nurs Health Sci 2014 Mar;16(1):126-130 [FREE Full text] [doi: 10.1111/nhs.12060] [Medline: 24118852]
9. World Health Organization. 2014. NCD country profile: Syrian Arab Republic URL: http://www.who.int/nmh/countries/syr_en.pdf [accessed 2017-03-22] [WebCite Cache ID 6pA0NpVy6]
10. World Health Organization. 2014. NCD country profile: Lebanon URL: http://www.who.int/nmh/countries/lbn_en.pdf [accessed 2017-03-22] [WebCite Cache ID 6pA0Qk1uI]
11. World Health Organization. 2015. Lebanon: WHO statistical profile URL: http://www.who.int/gho/countries/lbn.pdf?ua=1 [accessed 2017-08-23] [WebCite Cache ID 6swLN8RVJ]
12. World Health Organization. 2015. Syrian Arab Republic: WHO statistical profile URL: http://www.who.int/gho/countries/syr.pdf?ua=1 [accessed 2017-08-23] [WebCite Cache ID 6swLQRNPR]
13. World Bank. 2013 Sep 20. Lebanon economic and social impact assessment of the Syrian conflict URL: http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2013/09/24/000333037_20130924111238/Rendered/PDF/810980LB0box379831B00P14754500PUBLIC0.pdf [WebCite Cache ID 6pA0Way6y]
14. Khatib O. Noncommunicable diseases: risk factors and regional strategies for prevention and care. East Mediterr Health J 2004 Nov;10(6):778-788. [Medline: 16335764]
15. Crisp J, Morris T, Refstie H. Displacement in urban areas: new challenges, new partnerships. Disasters 2012 Jul;36(Suppl 1):S23-S42. [doi: 10.1111/j.1467-7717.2012.01284.x] [Medline: 22687155]
16. National Institute for Health and Care Excellence. NICE Guidance. 2011. Hypertension in adults: diagnosis and management URL: http://guidance.nice.org.uk/CG127 [WebCite Cache ID 6pA0bn6Hw]
17. World Health Professions Alliance. WHPA. 2017. Health improvement card URL: http://www.whpa.org/ncd_Health-Improvement-Card_web.pdf [accessed 2017-10-03] [WebCite Cache ID 6pA0gYuTD]
18. World Health Organization. 2010. Package of essential noncommunicable (PEN) disease interventions for primary health care in low-resource settings URL: http://apps.who.int/iris/bitstream/10665/44260/1/9789241598996_eng.pdf [WebCite Cache ID 6pA0i5Z7m]
19. Massachusetts Institute of Technology Sana mHealth Group. Sana at MIT. 2017. The patient controlled health record (PCHR) URL: http://sana.mit.edu/pchr/ [accessed 2017-03-22] [WebCite Cache ID 6pA0kkzHo]
20. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2008 Jan;31(Suppl 1):S55-S60. [doi: 10.2337/dc08-S055] [Medline: 18165338]
21. James PA, Oparil S, Carter BL, Cushman WC, Dennison-Himmelfarb C, Handler J, et al. 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). J Am Med Assoc 2014 Feb 5;311(5):507-520. [doi: 10.1001/jama.2013.284427] [Medline: 24352797]
22. Price M, Davies I, Rusk R, Lesperance M, Weber J. Applying STOPP guidelines in primary care through electronic medical record decision support: randomized control trial highlighting the importance of data quality. JMIR Med Inform 2017 Jun 15;5(2):e15 [FREE Full text] [doi: 10.2196/medinform.6226] [Medline: 28619704]
23. Steyn K, Lombard C, Gwebushe N, Fourie JM, Everett-Murphy K, Zwarenstein M, et al. Implementation of national guidelines, incorporated within structured diabetes and hypertension records at primary level care in Cape Town, South Africa: a randomised controlled trial. Glob Health Action 2013 Sep 25;6(1):20796. [doi: 10.3402/gha.v6i0.20796]
24. 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]
25. Adepoju IO, Albersen BJ, De Brouwere V, van Roosmalen J, Zweekhorst M. mHealth for clinical decision-making in sub-Saharan Africa: a scoping review. JMIR Mhealth Uhealth 2017 Mar 23;5(3):e38 [FREE Full text] [doi: 10.2196/mhealth.7185] [Medline: 28336504]
26. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Jul 15;16:572 [FREE Full text] [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]

27. Roehrs A, da Costa CA, Righi RD, de Oliveira KS. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
28. Reed M, Huang J, Graetz I. Use of electronic health record is associated with improved glycemic and cholesterol control for patients with diabetes. J Clin Outcomes Manag 2012;19(12).

Abbreviations

BMI: body mass index
BP: blood pressure
EHR: electronic health record
HbA1c: glycated hemoglobin
mHealth: mobile health
NCD: noncommunicable disease
PCHR: personally controlled health record
PHC: primary health care
UNHCR: United Nations High Commissioner for Refugees

Edited by G Eysenbach; submitted 23.03.17; peer-reviewed by F Pourmalek, C Exeter; comments to author 22.07.17; revised version received 23.08.17; accepted 10.09.17; published 18.10.17.

Please cite as:
Doocy S, Paik KE, Lyles E, Hei Tam H, Fahed Z, Winkler E, Kontunen K, Mkanna A, Burnham G
Guidelines and mHealth to Improve Quality of Hypertension and Type 2 Diabetes Care for Vulnerable Populations in Lebanon: Longitudinal Cohort Study
JMIR Mhealth Uhealth 2017;5(10):e158
URL: http://mhealth.jmir.org/2017/10/e158/
doi:10.2196/mhealth.7745
PMID:29046266

Original Paper

A Mobile Phone-Based Life Skills Training Program for Substance Use Prevention Among Adolescents: Pre-Post Study on the Acceptance and Potential Effectiveness of the Program, Ready4life

Severin Haug[1], PhD; Raquel Paz Castro[1], MSc; Christian Meyer[2], PhD; Andreas Filler[3,4], MSc; Tobias Kowatsch[3], PhD; Michael P Schaub[1], PhD

[1]Swiss Research Institute for Public Health and Addiction at the University of Zurich, Zurich University, Zurich, Switzerland

[2]Institute of Social Medicine and Prevention, University of Greifswald, Greifswald, Germany

[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[4]Energy Efficient Systems Group, University of Bamberg, Bamberg, Germany

Corresponding Author:
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction
University of Zurich
Konradstrasse 32
Zurich, 8031
Switzerland
Phone: 41 444481174
Email: severin.haug@isgf.uzh.ch

Abstract

Background: Substance use and misuse often first emerge during adolescence. Generic life skills training that is typically conducted within the school curriculum is effective at preventing the onset and escalation of substance use among adolescents. However, the dissemination of such programs is impeded by their large resource requirements in terms of personnel, money, and time. Life skills training provided via mobile phones might be a more economic and scalable approach, which additionally matches the lifestyle and communication habits of adolescents.

Objective: The aim of this study was to test the acceptance and initial effectiveness of an individually tailored mobile phone–based life skills training program in vocational school students.

Methods: The fully automated program, named *ready4life*, is based on social cognitive theory and addresses self-management skills, social skills, and substance use resistance skills. Program participants received up to 3 weekly text messages (short message service, SMS) over 6 months. Active program engagement was stimulated by interactive features such as quiz questions, message- and picture-contests, and integration of a friendly competition with prizes in which program users collected credits with each interaction. Generalized estimating equation (GEE) analyses were used to investigate for changes between baseline and 6-month follow-up in the following outcomes: perceived stress, self-management skills, social skills, at-risk alcohol use, tobacco smoking, and cannabis use.

Results: The program was tested in 118 school classes at 13 vocational schools in Switzerland. A total of 1067 students who owned a mobile phone and were not regular cigarette smokers were invited to participate in the life skills program. Of these, 877 (82.19%, 877/1067; mean age=17.4 years, standard deviation [SD]=2.7; 58.3% females) participated in the program and the associated study. A total of 43 students (4.9%, 43/877) withdrew their program participation during the intervention period. The mean number of interactive program activities that participants engaged in was 15.5 (SD 13.3) out of a total of 39 possible activities. Follow-up assessments were completed by 436 of the 877 (49.7%) participants. GEE analyses revealed decreased perceived stress (odds ratio, OR=0.93; 95% CI 0.87-0.99; *P*=.03) and increases in several life skills addressed between baseline and the follow-up assessment. The proportion of adolescents with at-risk alcohol use declined from 20.2% at baseline to 15.5% at follow-up (OR 0.70, 95% CI 0.53-0.93; *P*=.01), whereas no significant changes were obtained for tobacco (OR 0.94, 95% CI 0.65-1.36; *P*=.76) or cannabis use (OR 0.91, 95% CI 0.67-1.24; *P*=.54).

Conclusions: These results reveal high-level acceptance and promising effectiveness of this interventional approach, which could be easily and economically implemented. A reasonable next step would be to test the efficacy of this program within a controlled trial.

(JMIR Mhealth Uhealth 2017;5(10):e143) doi:10.2196/mhealth.8474

KEYWORDS

coping skills; social skills; substance use disorder; adolescents; students; mobile phone

Introduction

Several biological, psychological, and social transitions that occur during adolescence are essential for a young person's later-life trajectory [1,2]. These transitions offer opportunities for them to gain skills to achieve greater autonomy from adults, build social connections with peers, develop a positive body image, and form a sense of identity. However, these transitions also facilitate exploration and risk taking at a stage when cognitive functions of the brain are not yet fully developed [3]. Shifts of emotional regulations and increased risky behaviors result in vulnerabilities for mental and substance use disorders, which constitute the biggest contributors to the health burden of 10- to 24-year-old individuals [4]. Substance use and the development of substance use disorders often first emerge during adolescence and co-occur with mental disorders [1].

The age of onset of substance use is similar across high-income countries, with increasing levels and frequency of use beginning in mid-adolescence, peaking during early adulthood [5]. According to the World Health Organization (WHO) World Mental Health Surveys [6], the interquartile range of the age-of-onset distributions is typically 14 to 21 years for alcohol, 15 to 21 years for tobacco, 16 to 22 years for cannabis, and 19 to 28 years for cocaine.

As most young people do not fulfill criteria for problematic or disordered use, prevention programs and early intervention should be the focus, rather than substance-related treatment measures. A recent systematic review of studies assessing the effectiveness of prevention, early intervention, and harm reduction in young people for tobacco, alcohol, and illicit drugs demonstrated the effectiveness of taxation, public consumption bans, advertising restrictions, and minimum legal age, as well as the potential effectiveness of preventative interventions that deliver life skills training in educational settings [7]. Schools are particularly suitable settings to reach adolescents with preventative interventions because of the ease of delivery and access to young people within compulsory secondary education [7].

A Cochrane review on school-based programs for the prevention of tobacco smoking [8] concluded that combined social competence and social influence interventions had a significant effect at 1 year and at longest follow-up, whereas a social influences program on its own, multimodal community-wide initiatives, and information-only interventions were found to be ineffective. Another Cochrane review on school-based prevention programs for alcohol misuse in young people [9] concluded that generic psychosocial and developmental prevention programs can be effective. However, the methodological quality of the trials included in the analysis was poor, and this did not allow for any quantitative pooling of data. A Cochrane review on school-based prevention of illicit drug use [10] concluded that programs based on a combination of social competence and social influence approaches were most promising, on average exhibiting small but consistent protective effects to prevent drug use.

According to the WHO, life skills are "abilities for adaptive and positive behavior that enable individuals to deal effectively with the demands and challenges of everyday life" [11]. The majority of the generic programs addressing social competences and social influences that were included in the aforementioned reviews were based on Bandura's social learning theory [12], which hypothesizes that children and adolescents learn substance use by modeling, imitation, and reinforcement, influenced by individual cognitions, attitudes, and skills. Moreover, substance use susceptibility is increased by poor personal and social skills.

Generic life skills programs to prevent substance use, such as the IPSY (Information + Psychosocial Competence = Protection) program developed in Germany [13] or the ALERT [14] or LifeSkills Training [15] programs developed in the United States, typically combine training in self-management skills, social skills (eg, self-awareness, coping strategies, assertiveness, or communication skills), and substance use resistance skills (eg, resisting peer pressure to drink alcohol and recognizing and resisting media influences promoting cigarette smoking).

Although these life skills training programs were effective at preventing the onset of specific substances [8,13] or at decreasing problematic substance use [9], their implementation and dissemination in schools present serious challenges [16]. First, teachers and other professionals need the time, motivation, knowledge, and skills to deliver the program. Second, extensive resources—in terms of personnel, money, and time allocated to deliver substance use prevention—are required to prepare and administer such programs.

Electronically delivered interventions (eg, via computer, Internet, or mobile phone) have the potential to overcome the aforementioned obstacles that hinder successful program implementation and dissemination of life skills training in schools at a larger scale. Electronically delivered interventions have a wide reach at a low cost and offer the opportunity to automatically deliver individually tailored contents that can be accessed at any time and in any place [17]. Furthermore, electronically delivered interventions might be more appealing for adolescents because they can better ensure privacy and tailor contents to their needs.

Beyond traditional personal computers, a promising means of delivering prevention programs is to do so remotely through the use of mobile technologies. In Switzerland, as in most other developed countries, almost all (98%) adolescents between the ages of 12 and 19 years own a mobile phone, and 97% of these phones are smartphones [18]. Most adolescents are familiar with how to use mobile phones and typically use them on a daily basis for texting, taking pictures, playing games, and so on. Mobile phone–based interventions can provide almost constant support to users, relative to interventions that can only

be accessed at specific times or locations; and they provide a discrete and confidential means of intervention delivery [19].

Several studies have underlined the potential and effectiveness of substance-specific mobile phone–based programs for early interventions in adolescents already consuming specific substances such as tobacco or alcohol [20-22]. However, the feasibility and effectiveness of more generic, life skills interventions to prevent substance use via mobile phones have not been addressed to date.

Consequently, the objectives of this study were (1) to test the acceptance, use, and evaluation of a mobile phone–based life skills program among vocational school students and (2) to explore its potential effectiveness.

Methods

Setting

In most European countries, vocational schools are postsecondary public schools that are analogous to American community colleges. They are a part of the dual educational system that combines apprenticeships in a business context and vocational training in a school context. Vocational schools provide general education and specific skills for each particular profession.

On the basis of data from the Swiss Federal Statistical Office, approximately half of all Swiss adolescents aged 16 to 19 years currently attend vocational schools [23], with the highest proportions among adolescents aged 17 years (males: 60%, females: 48%) and 18 years (males: 58%, females: 47%).

Design and Procedures

A longitudinal pre-post study design with assessments at baseline and after program completion (month 6) was used to test the initial effectiveness of the program. Prevention specialists from branches of the Swiss Lung Association (Aargau, Basel-Land, Basel-Stadt, Berne, Vaud, and St Gallen), with particular training in the study and program to be delivered, arranged sessions lasting 30 min in participating vocational school classes during regular school lessons reserved for health education. Within this session, the students were informed about and invited to participate in a study testing innovative channels for the provision of health-related information and life skills. The students were informed by the prevention specialists about the study's aims and assessments, reimbursement, and data protection. Students were also informed that they could withdraw from program participation at any time, simply by sending a short service message (SMS) expressing their request to stop the program.

The mobile phone–based program and its association with a friendly competition with prizes were described in detail by the prevention specialists. To ensure sufficient participation and, thus, representativeness of the sample [24], students were informed that they would also receive a small reward for participating in the study. Each student was provided with a tablet computer or used his or her mobile phone for the screening process to assess for study eligibility and for study registration and the baseline assessment. Inclusion criteria for this study were (1) a minimum age of 16 years and (2) possession of a mobile phone. After being screened for the inclusion criteria and giving informed consent, study participants were invited to choose a username and provide their mobile phone number.

On the basis of previous results on the efficacy of texting-based programs for smoking cessation [25,26], vocational school students who smoked cigarettes regularly (at least four cigarettes over the preceding month and at least one cigarette within the preceding week) received a program version combining smoking cessation support based on the MobileCoach Tobacco program [27] and strategies for stress management. As this program primarily focused on smoking cessation and did not include comprehensive life skills training, we excluded regular smokers from this study.

After they had given their informed consent, study participants completed a baseline assessment directly on their mobile phone or on the tablet computer that they had been provided with. They received additional questions that were necessary to tailor their intervention's content. Subsequently, participants received individually tailored Web-based feedback directly on their mobile phone. Over the subsequent 6 months, they received individually tailored life skills training provided via mobile phone texting. All subjects were invited to complete a Web-based follow-up assessment after program completion 6 months after their enrollment in the study. For this, they received a text message (SMS) with a Web link to the Web-based assessment. Up to two text message–based reminders were sent to those who failed to complete the follow-up assessment upon initial request.

The study protocol was approved by the Ethics Committee in the Faculty of Philosophy at the University of Zurich, Switzerland (date of approval: June 24, 2016) and the trial conducted in compliance with the Declaration of Helsinki.

The Intervention Program ready4life

Theoretical Background and Intervention Contents

The intervention elements of the program called *ready4life* are based on social cognitive theory [28,29]. This theory relies on social learning theory, as it was founded on principles of learning within the human social context [12], though it has also integrated several concepts from cognitive psychology. Key concepts of this theory that are incorporated within the Web-based and text messaging–based life skills program are (1) outcome expectations (ie, beliefs about the likelihood and impact of the consequences of behavioral choices), (2) self-efficacy (ie, beliefs about one's personal ability to perform a desired behavior that could be stimulated; eg, by mastery, experience, or persuasion), (3) observational learning (ie, learning new behaviors via exposure to them through interpersonal or media displays; eg, through peer modeling), (4) facilitation (ie, providing strategies, tools, and resources that make new behaviors easier to perform), and (5) self-regulation (ie, controlling oneself via monitoring, goal setting, feedback, and self-instruction).

The contents of *ready4life* rely on proven and widely disseminated life skills programs such as IPSY [13], ALERT [14], and LifeSkills Training [15]. The program addresses (1)

self-management skills, (2) social skills, and (3) substance use resistance skills. As recruitment for this study was conducted within the German- and French-speaking part of Switzerland, all intervention contents were available in both German and French.

Technological Background

The intervention program was developed using the MobileCoach system. Technical details of the system are described elsewhere [30,31]. The MobileCoach system is available as an open-source project. Password protection and Secure Sockets Layer encoding are used to ensure the privacy and safety of data transfer.

Individually Tailored Feedback

Individually tailored Web-based feedback was given immediately after subjects completed the Web-based baseline assessment during school classes using tablet computers or mobile phones. This feedback comprised five screens, which included textual and graphical feedback on the following: (1) stress in general; (2) the individual level of stress in various domains; (3) the individual level of stress compared with an age- and gender-specific reference group, based on the Perceived Stress Scale (PSS) [32] and on data derived from a survey among Swiss vocational school students [33]; and (4) individually applied and suggested strategies to cope with stress.

Text Messages

For a period of 6 months, program participants received between two and four individualized text messages per week on their mobile phone. These messages were generated and sent by the fully automated MobileCoach system. Within the first 9 weeks, the messages focused on self-management skills; for example, coping with stress, emotional self-regulation, or management of feelings of anger and frustration. In the weeks 10 to 15, the messages focused on social skills, for example, making requests, refusing unreasonable requests, and meeting new people. In weeks 16 to 20, the text messages focused on substance use resistance skills, for example, recognizing and resisting media influences, social norms of substance use, or the associations of self-management skills and interpersonal competences with substance use. Boosters for each of the components were provided in weeks 22 to 24. The program concluded with information about a prize draw and an invitation to participate in the follow-up assessment. The messages were tailored according to the individual data from the baseline assessment and on-text messaging assessments during program runtime, for example, on substance use or the individual's emotional state. Sample messages from different intervention components are displayed in Figure 1.

To exploit the full potential of current mobile phones, several interactive features—such as quiz questions, tasks to create individually tailored if-then behavior plans based on implementation intentions, and message contests—were implemented within the program. Within picture and message contests, participants were invited to create and upload text messages or photos on specific topics, for example, on individually preferred strategies to cope with stress or on relaxation possibilities. The messages or pictures provided by other participants could be rated anonymously on a separate responsive website by all participants, and the top three postings were presented anonymously to all other participants after 48 hours on a website, which was only accessible for program participants. All messages and photos created by the participants were checked by a junior scientist with respect to the appropriateness of their content. Inappropriate content was excluded and not presented to the other participants.

Due to the wide dissemination of mobile phones in adolescents [18], several messages also included hyperlinks to audio files (eg, audio testimonials and motivational podcasts), as well as thematically appropriate video clips, pictures, and related websites.

Figure 1. Sample messages (translated from the German program version).

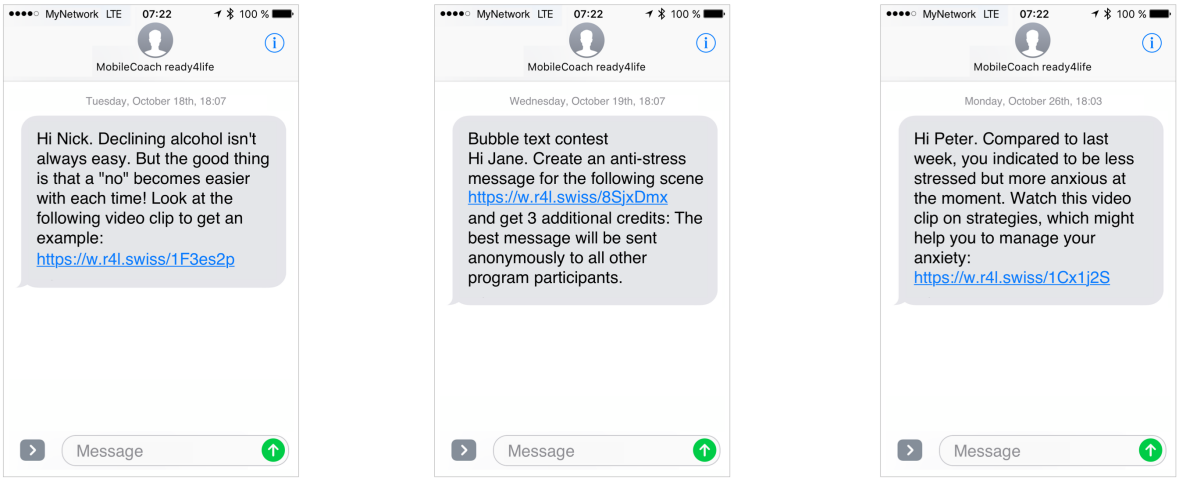
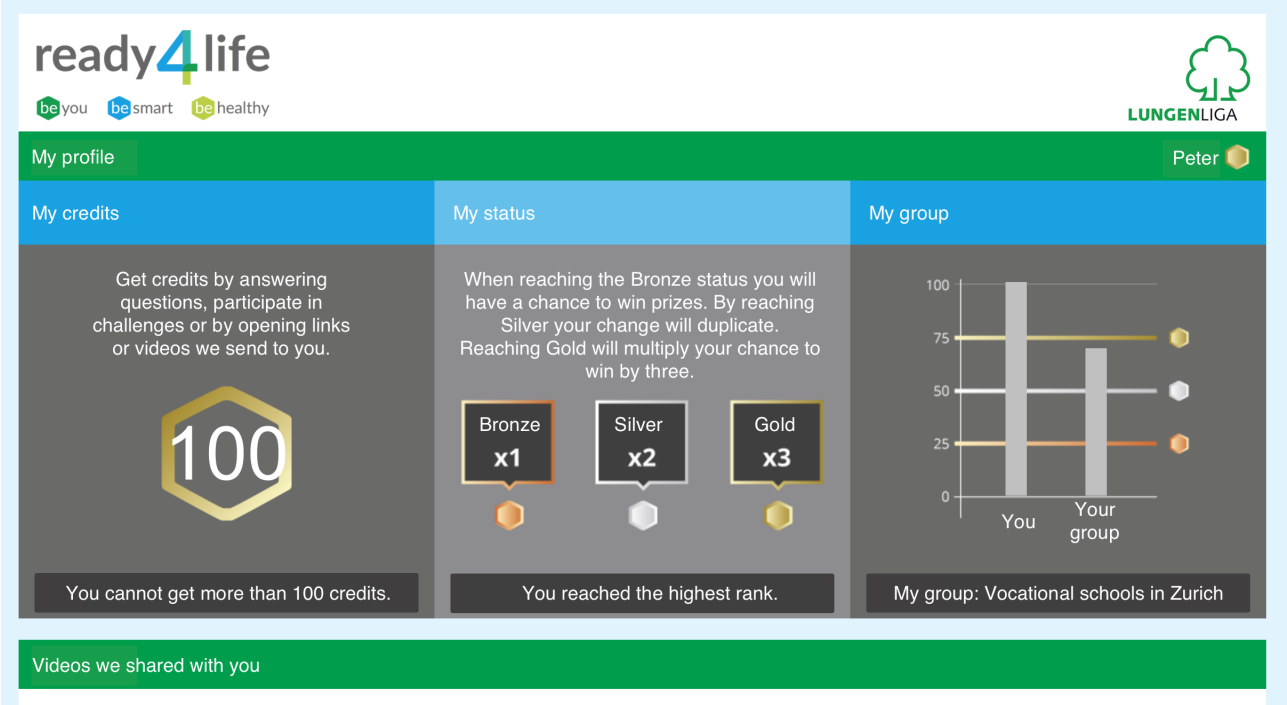


Figure 2. Individual profile page (translated from the German program version).



Prize Draw

To stimulate active program engagement, program use was associated with a friendly competition that allowed program users to collect credits for each interaction (eg, answering and monitoring text messages, participating in quizzes, creating messages or pictures within contests, and accessing video links integrated in text messages). The more credits participants collected, the higher were their chances of winning one of several attractive prizes that were part of a prize draw after program completion. Participants could retrieve their number of credits compared with the number of credits of other program participants of their group (similar starting date and Canton) at any time from an individual profile page (Figure 2).

Focus Groups for Pilot Testing

Before the study, a prototype of this program was tested and evaluated in five focus groups. Within these focus groups, individuals in the age range of 16 to 20 years, drawn from local vocational schools were asked to evaluate program flow, the layout and content of the Web-based assessment and feedback, and the content of text messages. Optimizations that resulted from these focus groups were integrated into the final version of the program.

Measures and Outcome Criteria

Demographics

The baseline survey included questions about the following demographic variables: gender, age, and migrant background. For the last of these, we asked about the country of birth of both parents of each vocational school student to identify potential migrant backgrounds. On the basis of this information, persons were assigned to one of the following categories: (1) persons with neither parent born outside Switzerland—no migrant background and (2) persons with one or both parents born outside Switzerland—migrant background.

Life Skills

Life skills were assessed at baseline and 6-month follow-up. The life skills assessed were related to the contents of the program and focused on (1) stress, (2) self-management skills, and (3) social skills.

A four-item version [32] of the PSS [34] was used to measure the degree to which students appraised situations as stressful over the preceding month. Responses were scored on a 1- to 5-point scale from (1) “never” to (5) “very often.” These four items were “In the last month...,” (1) “how often have you felt that you were unable to control the important things in your life?” (2) “how often have you felt confident about your ability to handle your personal problems?” (3) “how often have you felt that things were going your way?” and (4) “how often have you felt difficulties were piling up so high that you could not overcome them?” This scale showed comparable acceptable psychometric properties when administered online and in paper-pencil format, with a Cronbach alpha of .72 for the Web-based version [32].

Self-management and coping behavior within vocational training were assessed using one item derived from each of the five subscales of the Questionnaire for the Measurement of Stress and Coping in Children and Adolescents (SSKJ 3-8) [35]. The items were selected based upon their item-subscale correlation and the relevance of their content for vocational school students. Students indicated, on a 5-point rating scale ranging from never (1) to always (5), how often they use each of the presented coping strategies in response to stressful situations during vocational training. These five items were “If I am stressed during vocational training...,” (1) “I tell others, how I feel” (seeking social support); (2) “I change something so that things

are getting better" (problem solving); (3) "I tell myself that things will resolve themselves" (avoidant coping); (4) "I try to relax" (palliative emotion regulation); and (5) "I get totally upset" (anger-related emotion regulation). According to Eschenbeck et al [35], constructive coping behavior is particularly indicated by higher values on subscales (1), (2), and (4); conversely, higher values on subscales (3) and (5) are less desirable.

Social skills were assessed using a scale with seven items derived from the Assertion Inventory [36]. These items addressed (1) expressing an opinion that differs from that of other persons, (2) resisting social pressure to drink or smoke cigarettes, (3) telling a work colleague when he or she says or does something that bothers you, (4) asking questions to find out more about something, (5) apologizing when you are wrong, (6) accepting yourself even while being criticized, and (7) telling someone good news about yourself. Students indicated, on a 5-point rating scale ranging from never (1) to always (5), the frequency that they display each of the indicated behaviors. These seven items exhibited acceptable internal consistency, with a Cronbach alpha of .65.

Substance Use

Indicators of substance use were assessed at baseline and follow-up and included (1) at-risk alcohol use, (2) tobacco smoking, and (3) cannabis use.

At-risk alcohol use was assessed through the consumption items of the Alcohol Use Disorder Identification Test (AUDIT), the AUDIT-C [37]. The AUDIT-C assesses drinking quantity, drinking frequency, and binge drinking. On the basis of recommendations for adolescents [38], we used a cut-off value of ≥5 to determine whether risky drinking was present.

Tobacco smoking was assessed by the yes or no question: "have you taken at least one puff of a cigarette within the past 30 days?"

Cannabis use was assessed with the item "Within the last six months, how often did you use cannabis or marijuana?" with the response options (1) "never," (2) "1-5 times," (3) "6-20 times," and (4) "more often than 20 times." To estimate the prevalence of cannabis use within the last 6 months, we collapsed response options 2, 3, and 4 into a single category: cannabis use.

Program Use and Evaluation

To obtain the number of program participants who unsubscribed from the program within the program runtime of 6 months, we analyzed the log files of the MobileCoach system in which all incoming and outgoing text messages were recorded. Using these log files, we also assessed the mean number of replies to the 12 text message assessments during the program. At follow-up, we assessed another aspect of SMS usage by asking the participants whether they usually (1) read through the text messages thoroughly, (2) took only a short look at them, or (3) did not read the text messages.

Using a yes or no question, we evaluated whether the times when participants received the text messages were deemed to be appropriate. Furthermore, we assessed whether the number of received text messages was felt to be appropriate, or whether the participants would have preferred fewer or more messages. Finally, program participants were asked to rate the program and different program elements using the response categories "very good," "good," "less than good," "bad," and "don't know."

Outcome Criteria

To explore the intervention's effectiveness, the pre-post changes between baseline and 6-months follow-up of the following variables were investigated: (1) perceived stress [32], (2) self-management and coping behaviors [35], (3) interpersonal skills [36], (4) at-risk alcohol use [37], (5) tobacco smoking, and (6) cannabis use.

Data Analysis

To test for baseline differences between study participants and nonparticipants, Pearson χ^2analysis for categorical variables and nonpaired student's *t* tests for continuous variables were applied. For the attrition analysis (program participants lost to follow-up), we used χ^2analysis for categorical variables and *t* tests for continuous variables. Baseline equivalence and lack of attrition bias were assumed for tests with $P>.10$.

We used generalized estimating equation (GEE) analyses to investigate the longitudinal course of the outcome criteria over the study period of 6 months. GEE is a repeated-measures regression model that takes into account the correlation of repeated measures within each subject [32]. It is a powerful and versatile procedure for analyzing longitudinal data, with minimal assumptions about time dependence, and it allowed us to use all available longitudinal data, irrespective of single missing values at follow-up.

We used logistic GEE models for binary outcomes and linear GEE models for continuous variable outcomes. To control for attrition bias, we additionally added the respective baseline variables and variables on program use as covariates to the GEE models. Each GEE model included (1) the examined time variable (baseline vs follow-up assessment) as a predictor, (2) covariates to account for selective attrition, and (3) some outcome variable as the dependent variable.

Given the clustered nature of the data (students within school classes and intraclass correlation for the considered outcomes ranged from .02-.06), we computed robust variance estimators for all GEE analyses. An alpha level of .05 (two-tailed) was chosen for all statistical tests conducted in the study. All analyses were performed using the Stata software package, version 12 (StataCorp).

Results

Study Participants

Participants' progression through the study is depicted in Figure 3. At the time of the Web-based assessment in 118 vocational school classes at 13 Swiss vocational schools, a total of 2032 students were present. Among them, 1889 (92.96%, 1889/2032) had a minimum age of 16 years and owned a mobile phone and, as such, fulfilled the inclusion criteria for study participation.

A total of 822 (43.51%, 822/1889) regular tobacco smokers were not considered within this study because they received a program focusing on smoking cessation. Among the 1067 (56.48%, 1067/1889) students who were invited to participate in *ready4life*, 877 (82.19%, 877/1067) agreed to participate. Nonparticipants were subjects who met enrollment criteria but did not agree to participate in the study.

Table 1 summarizes characteristics of study participants and nonparticipants. Study participants had a mean age of 17.4 years (standard deviation, SD 2.7) and consisted of 58.3% females. Study participants differed from nonparticipants with respect to the baseline variables *gender* and *migrant background*. A greater percentage of study participants than nonparticipants were female (χ^2_1=5.5, *P*=.02) and had no migrant background (χ^2_1=4.8, *P*=.03).

Table 1. Baseline characteristics of study participants and nonparticipants.

Variable	Study participants (N=877)	Nonparticipants (N=190)	*P* value
Female gender, n (%)	511 (58.3)	93 (48.9)	.02
Age in years, mean (SD[a])	17.4 (2.7)	17.9 (4.5)	.10
Migrant background, n (%)			.03
No migrant background	460 (52.5)	83 (43.7)	
Migrant background	417 (47.5)	107 (56.3)	
Perceived stress (PSS[b], range 1-5), mean (SD)	2.51 (0.66)	2.53 (0.64)	.65
Self-management skills (range 1-5), mean (SD)			
Seeking social support	2.9 (1.3)	--	
Problem solving	3.3 (1.0)	--	
Avoidant coping	2.5 (1.1)	--	
Palliative emotion regulation	3.3 (1.2)	--	
Anger-related emotion regulation	2.6 (1.2)	--	
Social skills (scale, range 1-5), mean (SD)	3.7 (0.6)	--	
Alcohol use (AUDIT-C[c]), n (%)			.43
Not at risk (<5)	710 (81.0)	128 (83.7)[d]	
At risk (≥5)	167 (19.0)	25 (16.3)	
Tobacco smoking in the previous 30 days, n (%)			
No	796 (90.8)	--	
Yes	81 (9.2)	--	
Cannabis use in the previous 6 months, n (%)			.25
No	769 (87.7)	129 (84.3)[d]	
Yes	108 (12.3)	24 (15.7)	

[a]SD: standard deviation.

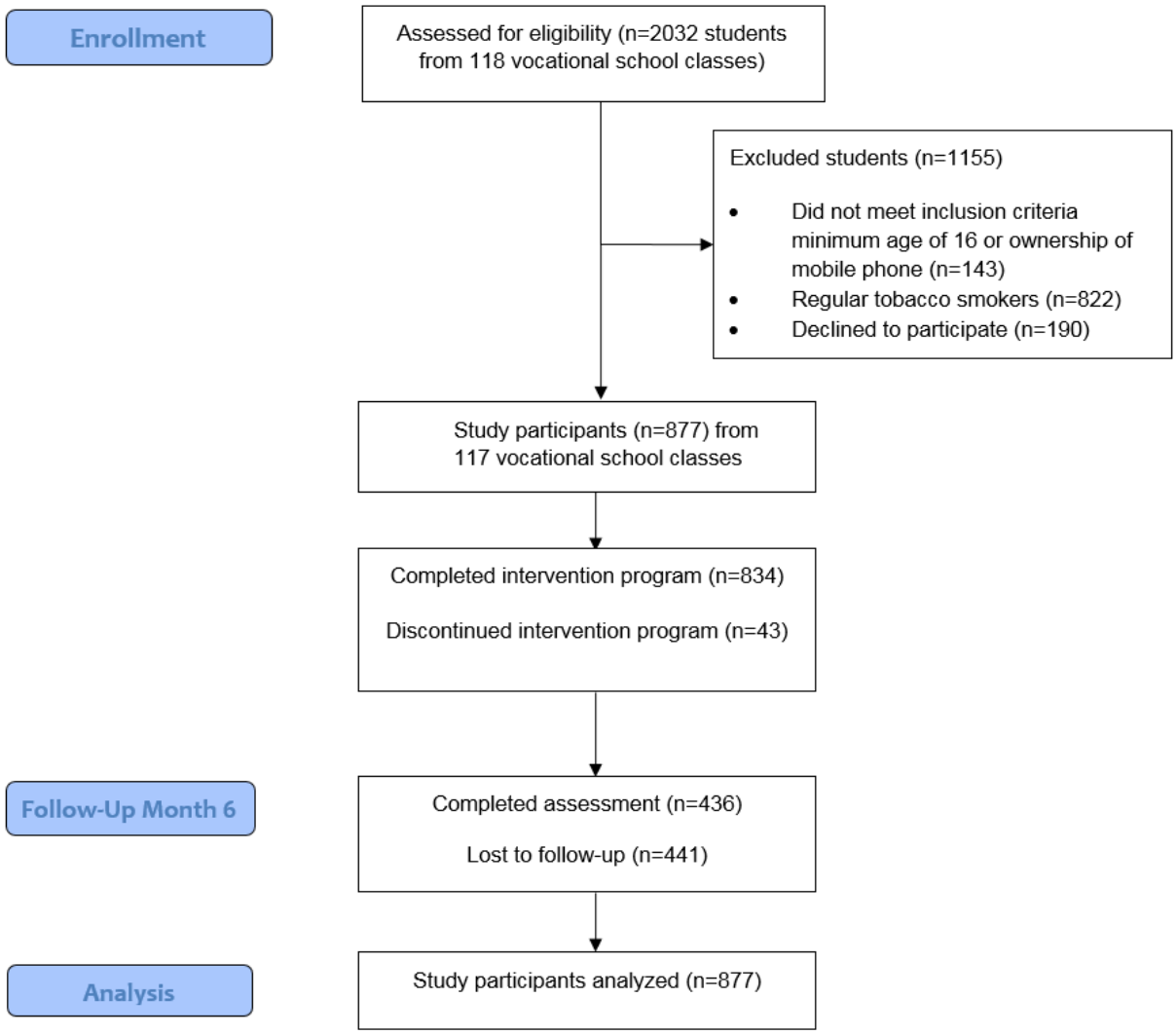
[b]PSS: Perceived Stress Scale.

[c]AUDIT-C: Alcohol Use Disorders Identification Test-C.

[d]n=37 missing values in nonparticipants.

Follow-up assessments were completed by 436 of the 877 (49.7%) study participants. Concerning attrition bias, the analysis revealed that follow-up assessments were completed more likely by female than male participants (χ^2_1=6.7, *P*<.01), by participants without a migrant background (χ^2_1=9.9, *P*<.01), by participants with higher values on the item on social support seeking (t_{875}=−2.47, *P*=.01), and by participants with a higher number of program activities during the program period (t_{875}=37.3, *P*<.001). To account for this attrition bias, these variables were entered as covariates within the GEE models that compared changes in outcomes between baseline and 6-month follow-up.

Figure 3. Participants' progress through the study.



Acceptability of the Intervention

During the program, which lasted for 6 months, 43 (4.9%) of the initial 877 program participants unsubscribed from the program, resulting in a total of 834 (95.1%, 834/877) participants who completed the entire program.

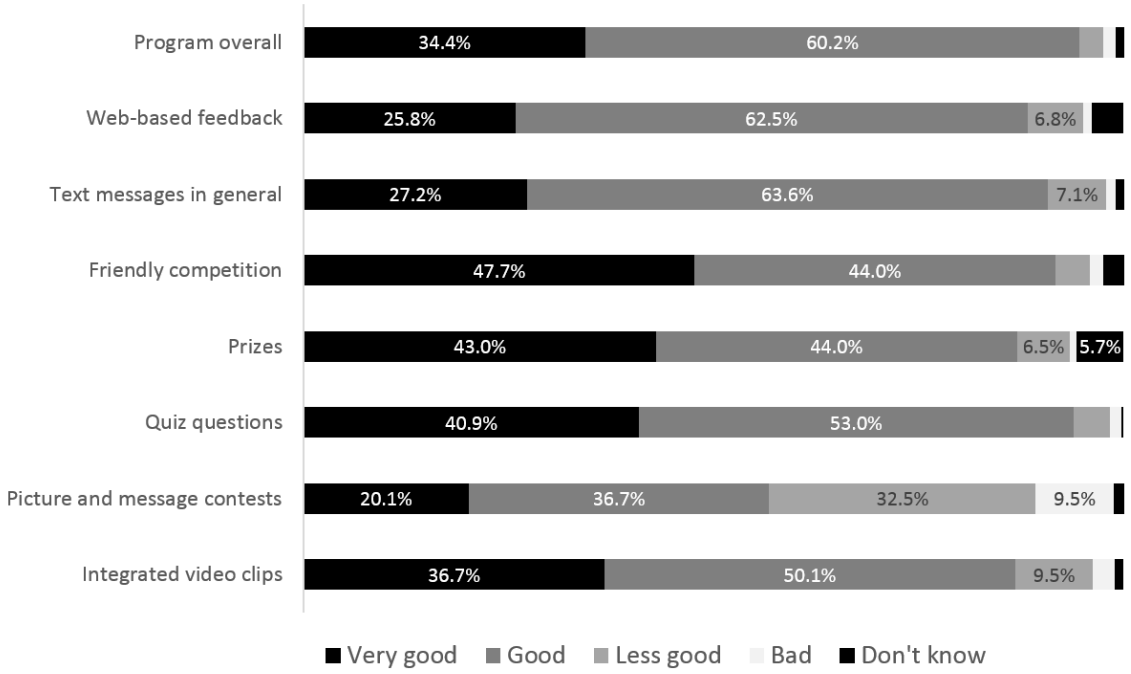
A total of 39 activities (eg, replies to texting prompts, accessing Web links within text messages, and participating in contests) were prompted over the 6-month program. The mean number of activities carried out by participants was 15.5 (SD 13.3). Of the 877 participants, 134 (15.3%) participated in no activities, 223 (25.4%) carried out 1 to 7 activities, 102 (11.6%) engaged in 8 to 14 activities, 82 (9.4%) in 15 to 21 activities, 123 (14.0%) in 22 to 28 activities, 156 (17.8%) in 29 to 35 activities, and 57 (6.5%) in 36 to 39 activities.

Of the 387 subjects with valid follow-up data, 323 (83.4%) indicated that they "read the SMS messages thoroughly," 61 persons (15.8%) reported that they "took a short look at the feedback messages," and only 3 persons (0.8%) chose the predefined response category "I did not read the feedback messages."

The duration of the program was rated as appropriate by 334 (86.7%, 334/385) program participants with valid follow-up data. The number of received SMS messages was rated as appropriate by 84.9% (328/386); 6.7% (26/386) would have preferred more, and 8.3% (32/386) would have preferred fewer SMS messages. Almost all participants reported that the text messages were comprehensible (98.7%, 379/384). Participants were also asked whether the text messages were helpful, and 296 out of 384 (77.1%) agreed with this. Three out of 4 participants (73.2%, 281/384) indicated that they perceived the text messages as individually tailored to them.

Figure 4 presents additional evaluations of the program and specific program elements. The program overall was evaluated as "very good" or "good" by 94.6% of the participants. Out of the specific program elements, the competition for prizes, the quiz questions, and the text messages in general received the best evaluations, with >90% of participants rating them "good" or "very good." The picture and message contests received the poorest ratings (56.8% "good" or "very good").

Figure 4. Evaluation of the program and specific program elements by program participants (n=384). Values are presented for percentages >5%.



Program Effectiveness

Life Skills

Pre-post comparisons of the variables addressing life skills are displayed in Table 2. The GEE analyses revealed a statistically significant decrease in perceived stress (odds ratio, OR 0.93, 95% CI 0.87-0.99; *P*=.03). Meanwhile, statistically significant increases were obtained for the items addressing the self-management skills *seeking social support* (OR 1.18, 95% CI 1.05-1.33; *P*=.008) and *palliative emotion regulation* (OR 1.13, 95% CI 1.01-1.28; *P*=.04), as well as for the scale addressing social skills (OR 1.07, 95% CI 1.00-1.13; *P*=.04).

Substance Use

Pre-post comparisons of substance use prevalence rates are displayed in Table 3. Concerning alcohol use, GEE analyses revealed a statistically significant decrease in the percentage of persons with at-risk alcohol use from the baseline assessment to the follow-up assessment (OR 0.70, 95% CI 0.53-0.93; *P*=.01). No significant pre-post differences were obtained in the percentage of persons using cannabis or smoking cigarettes.

Table 2. Pre-post comparisons of variables addressing life skills.

Variable	Pre mean (SD[a])	Post mean (SD)	OR[b](95% CI)[c] (N=877)	*P* value
Perceived stress (PSS[d], range 1-5)	2.5 (0.7)	2.4 (0.7)	0.93 (0.87-0.99)	.03
Self-management skills (range 1-5)				
Seeking social support	3.0 (1.2)	3.2 (1.2)	1.18 (1.05-1.33)	.008
Problem-solving	3.4 (1.0)	3.3 (1.0)	0.93 (0.83-1.04)	.23
Avoidant coping	2.4 (1.0)	2.5 (1.1)	1.09 (0.97-1.23)	.16
Palliative emotion regulation	3.2 (1.2)	3.4 (1.1)	1.13 (1.01-1.28)	.04
Anger-related emotion regulation	2.7 (1.2)	2.6 (1.2)	0.99 (0.87-1.11)	.81
Social skills (range 1-5)	3.8 (0.6)	3.9 (0.6)	1.07 (1.00-1.13)	.04

[a]SD: standard deviation.

[b]OR: odds ratio.

[c]Linear generalized estimation equation (GEE) models, with time variable (baseline vs follow-up assessment) as the predictor, adjusted for attrition bias.

[d]PSS: Perceived Stress Scale.

Table 3. Pre-post comparisons of variables addressing substance use.

Variable	Pre n (%)	Post n (%)	OR[a](95% CI)[b] (N=877)	*P* value
At-risk alcohol use, AUDIT-C[c](N=420)	85 (20.2)	65 (15.5)	0.70 (0.53-0.93)	.01
Tobacco smoking, previous 30 days (N=392)	33 (8.4)	31 (7.9)	0.94 (0.65-1.36)	.76
Cannabis use, previous 6 months (N=419)	44 (10.5)	40 (9.5)	0.91 (0.67-1.24)	.54

[a]OR: odds ratio.

[b]Logistic generalized estimation equation (GEE) models, with time variable (baseline vs follow-up assessment) as predictor, adjusted for attrition bias.

[c]AUDIT-C: Alcohol Use Disorders Identification Test-C.

Discussion

Principal Findings

Within this study, we tested the acceptability and explored the potential effectiveness of a newly developed mobile phone–based life skills training program for substance use prevention among adolescents. The study revealed three main findings: (1) concerning participation, a large proportion of the eligible adolescents who were invited for program and study participation in the setting of a school classroom, participated; (2) concerning program use, the majority of program participants completed the entire program and engaged in program activities; however, regular program use could be improved; and (3) concerning program effectiveness, the initial results derived from this pre-post comparison revealed statistically significant increases in the life skills addressed, a decline in at-risk alcohol use, and stable prevalence rates for tobacco and cannabis use.

The proactive invitation for program and study participation in the school setting, in combination with the offer of a low-threshold mobile phone–based intervention, permitted us to reach 4 out of 5 adolescents for participation in the life skills program *ready4life*. Given the program duration of 6 months and that program participants needed to indicate their mobile phone number, this high participation rate is particularly remarkable and was even higher than for substance-specific mobile phone–based programs conducted in the same setting and using similar recruitment procedures; between 50% and 75% participated in comparable programs to support smoking cessation [25,26], whereas 75% participated in comparable programs to reduce problem drinking [39,40]. Beyond proactive recruitment in the school setting and during school hours, the following reasons might have contributed to the high participation rate we observed: (1) adolescents were invited by an institution independent of their school and teacher (anonymity); (2) the mobile phone–based program was flexible for use at any time and in any place, and withdrawal from the program was permitted at any time; (3) program participation and use were associated with participation in a friendly competition with the chance to win one of several attractive prizes; and (4) the program contents were developed specifically for adolescents during their vocational training.

Participation in the program was lower in male adolescents and among those reporting an immigrant background. These findings should be considered for program optimization, for example, by highlighting the relevance of this program for these subgroups or by emphasizing interesting program elements, focusing particularly on these target groups, during program presentations in school classrooms.

Overall acceptance of the intervention was good. Nearly all program participants (95%) stayed logged in until the end of the program, which lasted 6 months. The SMS messages were read by almost all program participants (94%), and 3 out of 4 participants reported that they were helpful and perceived the text messages as individually tailored to them. However, 15% failed to engage in any of the 39 program activities, and 52% engaged in fewer than half of the possible activities. On the basis of this finding, there is clearly room for improvement in terms of active program engagement, particularly concerning the picture and message contests, which received the poorest ratings among all program elements. The poor rating for this highly interactive element might be because of the limitations of mobile phone texting to receive and send pictures, which could be implemented more elegantly within a chat-based native mobile phone app.

The results concerning the initial effectiveness of this program derived from a pre-post investigation are promising. Data revealed a decrease in perceived stress, an increase in social skills, and increases in two out of the three desirable self-management strategies (seeking social support and palliative emotion regulation), whereas no changes were observed in less desirable self-management strategies (avoidant coping and anger-related emotion regulation). The proportion of adolescents with at-risk alcohol use was reduced by a quarter from baseline assessment to follow-up, whereas no significant changes were obtained in the prevalence of tobacco and cannabis use. On the basis of typically increasing levels and frequency of substance use in adolescence and early adulthood [41], these stable or decreasing prevalence rates might be attributable to program participation. However, no final conclusions on program effectiveness should be drawn from this study, as we could not control other factors such as fluctuations that might have occurred over the course of the year.

Limitations

Beyond the limitations associated with the pre-post study design, some other study limitations should be mentioned. First, the results are restricted to adolescents without regular cigarette use, as only they were deemed eligible to participate in this general mobile phone–based life skills training. Second, only 50% of the study participants completed the follow-up assessment, which might have biased evaluations of the program

and the results on efficacy, even though we controlled for attrition bias in our GEE models. Third, since the study focused on program appropriateness, meaning that we wanted it to be relevant to actual prevention practices, we restricted our outcome assessments to either short forms of, or single items extracted from more extensive validated instruments; furthermore all outcomes were self-reported.

Conclusions and Outlook

This is the first study to test a comprehensive life skills training program for substance use prevention delivered by a mobile phone among adolescents. Our results suggest that this program, which delivers individualized messages and interactive activities integrated within a friendly competition, is both appropriate and promising in its effectiveness. Moreover, this intervention could be easily and economically implemented. On the basis of these initial positive results, a reasonable next step would be to test the efficacy of this program within a controlled trial.

Beyond testing the efficacy of digital-delivered life skills training programs, the examination of moderators and mediators of life skills trainings outcome in general remains an interesting question, which should be addressed in future studies.

Concerning moderators, it would be of particular interest to examine whether individuals with higher levels of substance use could also benefit from life skills training programs and to compare the effectiveness of these general life skills training programs with substance-specific interventions, for example, mobile phone–based programs for adolescents already consuming specific substances such as tobacco or alcohol [20-22]. Results from the IPSY program, conducted in young adolescents from Germany [13], indicated that this school-based and face-to-face delivered life skills training was ineffective for adolescents who are on a problematic developmental pathway of alcohol use. The authors conclude that this subgroup might be in need of an earlier, more intensive and tailored treatment compared with IPSY. In contrast, findings from the ALERT Plus project delivered in seventh and eighth grade, with booster lessons in the ninth grade, showed that curricula during high school can also be effective with at-risk youth, who are particularly likely to escalate drug use and experience drug-related harms [42].

Concerning mediators, it would be of particular interest to test which of the life skills addressed and successfully modified in turn might prevent or decrease problematic substance use, for example, results from the ALERT Plus project showed that program-induced changes in perceived social influences, one's ability to resist those influences, and beliefs about the consequences of drug use mediated the effects on drug use [42].

Another interesting topic which should be addressed in future studies concerns potential spillover effects of general life skills training programs on other mental health conditions. As substance use and other mental disorders, for example, alcohol use disorders and depression, do not emerge as single impairments but rather co-occur [1,43], effects of these programs on further mental health–related outcomes should also be investigated. Concerning the comparison of substance-specific interventions and more general life skills training programs, one could assume that the latter might have an impact on a wider spectrum of mental health conditions.

Acknowledgments

Funding for this study was provided by the Lung Association. The authors would like to thank all the involved prevention specialists within the Cantonal Lung Leagues in Aargau, Basel-Land, Basel-Stadt, Berne, Vaud, and St Gallen. Finally, they would like to thank all the vocational school teachers and students who supported and participated in the project.

Conflicts of Interest

Single authors (SH, RPC, and AF) were also involved in the development of the intervention. The funding institution did not influence the design and conduct of the study; the management, analysis, or interpretation of data; or the preparation, review, or approval of the manuscript.

References

1. Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry 2011 Oct;52(10):1015-1025 [FREE Full text] [doi: 10.1111/j.1469-7610.2011.02446.x] [Medline: 21815892]
2. Hall WD, Patton G, Stockings E, Weier M, Lynskey M, Morley KI, et al. Why young people's substance use matters for global health. Lancet Psychiatry 2016;3(3):265-279. [doi: 10.1016/S2215-0366(16)00013-44406229] [Medline: 26905482]
3. White AM. Understanding adolescent brain development and its implications for the clinician. Adolesc Med State Art Rev 2009 Apr;20(1):73-90, viii-ix. [Medline: 19492692]
4. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
5. Degenhardt L, Stockings E, Patton G, Hall WD, Lynskey M. The increasing global health priority of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):251-264. [doi: 10.1016/S2215-0366(15)00508-8] [Medline: 26905480]

6. Degenhardt L, Chiu WT, Sampson N, Kessler RC, Anthony JC, Angermeyer M, et al. Toward a global view of alcohol, tobacco, cannabis, and cocaine use: findings from the WHO World Mental Health Surveys. PLoS Med 2008 Jul 01;5(7):e141 [FREE Full text] [doi: 10.1371/journal.pmed.0050141] [Medline: 18597549]
7. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/S2215-0366(16)00002-X] [Medline: 26905481]
8. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. Cochrane Database Syst Rev 2013 Apr 30(4):CD001293. [doi: 10.1002/14651858.CD001293.pub3] [Medline: 23633306]
9. Foxcroft DR, Tsertsvadze A. Universal school-based prevention programs for alcohol misuse in young people. Cochrane Database Syst Rev 2011 May 11(5):CD009113. [doi: 10.1002/14651858.CD009113] [Medline: 21563171]
10. Faggiano F, Minozzi S, Versino E, Buscemi D. Universal school-based prevention for illicit drug use. Cochrane Database Syst Rev 2014(12):CD003020. [doi: 10.1002/14651858.CD003020.pub3] [Medline: 25435250]
11. World Health Organization. asksource.info. Geneva: World Health Organization; 1994. Life skills education for children and adolescents in schools URL: http://www.asksource.info/pdf/31181_lifeskillsed_1994.pdf [accessed 2017-09-12] [WebCite Cache ID 6tPyngAqM]
12. Bandura A. Social learning theory of aggression. J Commun 1978;28(3):12-29. [Medline: 690254]
13. Weichold K, Blumenthal A. Long-term effects of the life skills program IPSY on substance use: results of a 4.5-year longitudinal study. Prev Sci 2016 Jan;17(1):13-23. [doi: 10.1007/s11121-015-0576-5] [Medline: 26202801]
14. Ellickson PL, McCaffrey DF, Ghosh-Dastidar B, Longshore DL. New inroads in preventing adolescent drug use: results from a large-scale trial of project ALERT in middle schools. Am J Public Health 2003 Nov;93(11):1830-1836. [Medline: 14600049]
15. Botvin GJ, Griffin KW. Life skills training: preventing substance misuse by enhancing individual and social competence. In: Weichold K, Giannotta F, editors. Theory-Based Approaches to Substance Misuse and Abuse Prevention in School: New Directions for Youth Development. San Francisco: Jossey-Bass; 2014:57-66.
16. Ellickson PL. You've shown the program is effective. Now what? In: Weichold K, Giannotta F, editors. Theory-Based Approaches to Substance Misuse and Abuse Prevention in School: New Directions for Youth Development. San Francisco: Jossey-Bass; 2014:95-106.
17. Kuntsche E, Labhart F. The future is now--using personal cellphones to gather data on substance use and related factors. Addiction 2014 Jul;109(7):1052-1053. [doi: 10.1111/add.12540] [Medline: 24754712]
18. Waller G, Willemse I, Genner S, Suter L, Süss D. Swisscom. Zurich: Zürcher Hochschule für angewandte Wissenschaften; 2016. JAMES–Jugend, Aktivitäten, Medien–Erhebung Schweiz-Ergebnisbericht zur JAMES-Studie 2016 URL: https://www.swisscom.ch/content/dam/swisscom/de/ghq/verantwortung/documents/james2016/ergebnisbericht_james2016de.pdf [accessed 2017-09-12] [WebCite Cache ID 6tQ07fOwK]
19. Kenny R, Dooley B, Fitzgerald A. Feasibility of "CopeSmart": a telemental health app for adolescents. JMIR Ment Health 2015 Aug 10;2(3):e22 [FREE Full text] [doi: 10.2196/mental.4370] [Medline: 26552425]
20. Marsch LA, Borodovsky JT. Technology-based interventions for preventing and treating substance use among youth. Child Adolesc Psychiatr Clin N Am 2016 Oct;25(4):755-768. [doi: 10.1016/j.chc.2016.06.005] [Medline: 27613350]
21. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328. [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
22. Suffoletto B. Text message behavioral interventions: from here to where? Curr Opin Psychol 2016 Feb 1;9:16-21. [doi: 10.1016/j.copsyc.2015.09.012] [Medline: 26665157]
23. bfs.admin. Neuchâtel: Bundesamt für Statistik; 2017. Schulbesuchsquoten der 16-26-Jährigen 2015/16 [school enrollment rates of 16 to 26 year olds in 2015/16] URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft.assetdetail.2242832.html [accessed 2017-09-12] [WebCite Cache ID 6tQ0dk81g]
24. Edwards P, Cooper R, Roberts I, Frost C. Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires. J Epidemiol Community Health 2005 Nov;59(11):987-999 [FREE Full text] [doi: 10.1136/jech.2005.034397] [Medline: 16234429]
25. Haug S, Bitter G, Hanke M, Ulbricht S, Meyer C, John U. [Brief text messaging (SMS)-delivered smoking cessation intervention for apprentices in vocational schools: a longitudinal intervention study for process and outcome evaluation]. Gesundheitswesen 2013 Oct;75(10):625-631. [doi: 10.1055/s-0032-1331728] [Medline: 23292965]
26. Haug S, Schaub MP, Venzin V, Meyer C, John U. Efficacy of a text message-based smoking cessation intervention for young people: a cluster randomized controlled trial. J Med Internet Res 2013;15(8):e171 [FREE Full text] [doi: 10.2196/jmir.2636] [Medline: 23956024]
27. Haug S, Castro RP, Filler A, Kowatsch T, Fleisch E, Schaub MP. Efficacy of an Internet and SMS-based integrated smoking cessation and alcohol intervention for smoking cessation in young people: study protocol of a two-arm cluster randomised controlled trial. BMC Public Health 2014;14:1140 [FREE Full text] [doi: 10.1186/1471-2458-14-1140] [Medline: 25369857]
28. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

29. McAlister AL, Perry CL, Parcel GS. How individuals, environments, and health behaviour interact: social cognitive theory. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco: Jossey-Bass; 2008.
30. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: overview and preliminary evaluation in the public health context. 2015 Presented at: Wireless Telecommunications Symposium 2015 (WTS 2015); April 15-17, 2015; New York. [doi: 10.1109/WTS.2015.7117255]
31. Haug S, Kowatsch T, Castro RP, Filler A, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in young people: study protocol of a cluster-randomised controlled trial. BMC Public Health 2014 Aug 07;14:809 [FREE Full text] [doi: 10.1186/1471-2458-14-809] [Medline: 25099872]
32. Herrero J, Meneses J. Short web-based versions of the perceived stress (PSS) and Center for Epidemiological Studies-Depression (CESD) scales: a comparison to pencil and paper responses among Internet users. Comput Human Behav 2006 Sep;22(5):830-846. [doi: 10.1016/j.chb.2004.03.007]
33. Liakoni E, Schaub MP, Maier LJ, Glauser GV, Liechti ME. The use of prescription drugs, recreational drugs, and "soft enhancers" for cognitive enhancement among swiss secondary school students. PLoS One 2015;10(10):e0141289 [FREE Full text] [doi: 10.1371/journal.pone.0141289] [Medline: 26505633]
34. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
35. Eschenbeck H, Kohlmann CW, Lohaus A, Klein-Heßling J. The assessment of coping with the "Questionnaire for the Measurement Stress and Coping in Children and Adolescents (SSKJ 3-8)": factord and psychometric analyses. Diagnostica 2006 Jul;52(3):131-142. [doi: 10.1026/0012-1924.52.3.131]
36. Gambrill ED, Richey CA. An assertion inventory for use in assessment and research. Behav Ther 1975 Jul;6(4):550-561. [doi: 10.1016/S0005-7894(75)80013-X]
37. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol use disorders identification test. Arch Intern Med 1998;158(16):1789-1795. [Medline: 9738608]
38. Rumpf H, Wohlert T, Freyer-Adam J, Grothues J, Bischof G. Screening questionnaires for problem drinking in adolescents: performance of AUDIT, AUDIT-C, CRAFFT and POSIT. Eur Addict Res 2013;19(3):121-127. [doi: 10.1159/000342331] [Medline: 23183686]
39. Haug S, Castro RP, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
40. Haug S, Castro RP, Schaub MP. Erreichbarkeit jugendlicher für ein Internet- und SMS-programm zum thema alkohol [Accessibility of adolescents for an Internet- and text message-based program concerning alcohol abuse]. Sucht 2015;61(6):357-365. [doi: 10.1024/0939-5911.a000395]
41. Degenhardt L, Ferrari AJ, Calabria B, Hall WD, Norman RE, McGrath J, et al. The global epidemiology and contribution of cannabis use and dependence to the global burden of disease: results from the GBD 2010 study. PLoS One 2013;8(10):e76635 [FREE Full text] [doi: 10.1371/journal.pone.0076635] [Medline: 24204649]
42. Longshore D, Ellickson PL, McCaffrey DF, St Clair PA. School-based drug prevention among at-risk adolescents: effects of ALERT plus. Health Educ Behav 2007;34(4):651-668. [Medline: 17567823]
43. Fergusson DM, Boden JM, Horwood LJ. Tests of causal links between alcohol abuse or dependence and major depression. Arch Gen Psychiatry 2009;66(3):260-266. [Medline: 19255375]

Abbreviations

AUDIT-C: Alcohol Use Disorder Identification Test-C
GEE: generalized estimating equation
IPSY: Information + Psychosocial Competence = Protection
OR: odds ratio
PSS: Perceived Stress Scale
SD: standard deviation
SMS: short message service
WHO: World Health Organization

Edited by G Eysenbach; submitted 14.07.17; peer-reviewed by R Schwarzer, B Suffoletto; comments to author 02.08.17; revised version received 10.08.17; accepted 10.08.17; published 04.10.17.

Please cite as:
Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub MP
A Mobile Phone-Based Life Skills Training Program for Substance Use Prevention Among Adolescents: Pre-Post Study on the Acceptance and Potential Effectiveness of the Program, Ready4life
JMIR Mhealth Uhealth 2017;5(10):e143
URL: https://mhealth.jmir.org/2017/10/e143/
doi:10.2196/mhealth.8474
PMID:28978498

Original Paper

User Participation and Engagement With the See Me Smoke-Free mHealth App: Prospective Feasibility Trial

Chris A Schmidt[1*], PhD; James K Romine[1*], MPH; Melanie L Bell[1], PhD; Julie Armin[2], PhD; Judith S Gordon[3], PhD

[1]College of Public Health, University of Arizona, Tucson, AZ, United States

[2]Family and Community Medicine, University of Arizona, Tucson, AZ, United States

[3]College of Nursing, University of Arizona, Tucson, AZ, United States

[*]these authors contributed equally

Corresponding Author:
Judith S Gordon, PhD
College of Nursing
University of Arizona
1305 N. Martin Avenue
Tucson, AZ, 85721
United States
Phone: 1 520 626 4970
Email: judithg@email.arizona.edu

Abstract

Background: The See Me Smoke-Free (SMSF) mobile health (mHealth) app was developed to help women quit smoking by targeting concerns about body weight, body image, and self-efficacy through cognitive behavioral techniques and guided imagery audio files addressing smoking, diet, and physical activity. A feasibility trial found associations between SMSF usage and positive treatment outcomes. This paper reports a detailed exploration of program use among eligible individuals consenting to study participation and completing the baseline survey (*participants*) and ineligible or nonconsenting app installers (*nonparticipants*), as well as the relationship between program use and treatment outcomes.

Objective: The aim of this study was to determine whether (1) participants were more likely to set quit dates, be current smokers, and report higher levels of smoking at baseline than nonparticipants; (2) participants opened the app and listened to audio files more frequently than nonparticipants; and (3) participants with more app usage had a higher likelihood of self-reported smoking abstinence at follow up.

Methods: The SMSF feasibility trial was a single arm, within-subjects, prospective cohort study with assessments at baseline and 30 and 90 days post enrollment. The SMSF app was deployed on the Google Play Store for download, and basic profile characteristics were obtained for all app installers. Additional variables were assessed for study participants. Participants were prompted to use the app daily during study participation. Crude differences in baseline characteristics between trial participants and nonparticipants were evaluated using *t* tests (continuous variables) and Fisher exact tests (categorical variables). Exact Poisson tests were used to assess group-level differences in mean usage rates over the full study period using aggregate Google Analytics data on participation and usage. Negative binomial regression models were used to estimate associations of app usage with participant baseline characteristics after adjustment for putative confounders. Associations between app usage and self-reported smoking abstinence were assessed using separate logistic regression models for each outcome measure.

Results: Participants (n=151) were more likely than nonparticipants (n=96) to report female gender (P<.02) and smoking in the 30 days before enrollment (P<.001). Participants and nonparticipants opened the app and updated quit dates at the same average rate (rate ratio [RR] 0.98; 95% CI 0.92-1.04; P=.43), but participants started audio files (RR 1.07; 95% CI 1.00-1.13; P<.04) and completed audio files (RR 1.11; 95% CI 1.03-1.18; P<.003) at significantly higher rates than nonparticipants. Higher app usage among participants was positively associated with some smoking cessation outcomes.

Conclusions: This study suggests potential efficacy of the SMSF app, as increased usage was generally associated with higher self-reported smoking abstinence. A planned randomized controlled trial will assess the SMSF app's efficacy as an intervention tool to help women quit smoking.

(JMIR Mhealth Uhealth 2017;5(10):e142) doi:10.2196/mhealth.7900

KEYWORDS

mHealth; mobile health; intervention; smoking; diet; physical activity

Introduction

Mobile technology offers new tools for delivering cost-effective and scalable health interventions to diverse populations [1]. Smoking cessation is an important focus for mobile health (mHealth) initiatives, considering the heavy burden of tobacco-related morbidity and mortality [2] and the exponential increase in access to mobile phone devices [3]. A recent meta-analysis suggests that mHealth cessation programs result in increased quit rates among participants [4], although the degree of usage needed to attain desired health outcomes is often uncertain [5]. Furthermore, the degree of usage may be influenced by sociodemographic factors, age, psychological traits, and other user characteristics [6]. These factors should be considered when evaluating the effectiveness of mHealth interventions, such as the recently-developed See Me Smoke-Free (SMSF) multi-behavioral smoking cessation program.

The SMSF mHealth app was developed to help women quit smoking by targeting concerns about body weight, body image, and self-efficacy through the use of guided imagery audio files that address smoking, diet, and exercise [7]. The development of SMSF has been reported elsewhere [8]. Gordon et al [9] assessed and presented feasibility outcomes and exploratory analyses of program impact and identified possible associations between app usage and both smoking and dietary behavioral change. As program use was associated with positive treatment outcomes in the pilot study [9], further exploration of app use data was warranted. This study expands our understanding of feasibility outcomes of SMSF by exploring the representativeness of our sample versus all users of the app. As others have noted, analysis of utilization by user characteristics may have implications for program enhancements and dissemination of the app to appropriate populations [10]. Moreover, considering the broad array of smoking cessation programs on the market, feature-level analyses may point to elements that predict desired outcomes such as smoking abstinence [11].

SMSF was deployed to the Google Play Store and available for use by the public. All those who downloaded the app were invited to enroll in the study, and approximately half of all users enrolled as participants [9]. We collected profile data and Google Analytics usage data for all users and outcome data for study participants. *Participants* were defined as eligible individuals who consented to study participation and completed the baseline survey, whereas *nonparticipants* were ineligible or nonconsenting app installers. In this paper, we report on our analysis of these data to explore program use among all those who downloaded the app, as well as the relationship between program use and treatment outcomes. Our goals were to (1) compare baseline characteristics of SMSF trial participants with people who downloaded and used the app but did not participate in the study (nonparticipants), (2) estimate associations between participant baseline characteristics and app usage, (3) evaluate whether trial participation is associated with higher app usage, and (4) assess associations between app usage and smoking cessation among participants. We hypothesized that the participants are more likely to set quit dates, be current smokers, and report higher levels of smoking at baseline than nonparticipants; that participants open the app and listen to audio files more frequently than nonparticipants; and that participants with more app usage have higher likelihood of smoking cessation at follow-up.

We were interested in exploring potential differences between participants and nonparticipants for two main reasons, which are as follows: (1) to understand how app users who enroll in a research study compare with those who do not and (2) to determine whether those who did not meet our initial criteria would still use the app and find it potentially useful. Although SMSF was not marketed to nonsmokers, it is possible that individuals who had stopped smoking for 30 days or more might use the program to prevent relapse. It is also possible that nonsmoking individuals may have downloaded the app to consider whether it was something that they would want to recommend to a friend or family member who smoked.

Reporting follows the consolidated standards of reporting trials (CONSORT) guidelines for feasibility trials [12].

Methods

Study Design

The SMSF feasibility trial was a single arm, within-subjects, prospective cohort study (ClinicalTrials.gov NCT02972515). Details of this trial, including eligibility criteria, have been described elsewhere [7]. Briefly, the SMSF app delivered five imagery audio files, one of which was a general introduction to guided imagery [9]. Three files were designed to target specific behaviors—smoking cessation, physical activity, and fruit and vegetable consumption—and another addressed general wellness, positive body image, and self-efficacy [9]. Study participants were recruited via news stories and social media postings. Individuals who installed the app and completed app setup during the recruitment period (April 1, 2015 to July 31, 2015) were invited to participate in the feasibility study. Individuals were eligible for participation if they identified as female, were at least 18 years old, smoked in the last 30 days, lived in the United States, used an Android phone, agreed to use the app "most days for 30 days," spoke English, and had a valid email address [7]. After eligibility screening, eligible individuals consenting to study participation and completing the baseline survey were enrolled (*participants*). Both study participants and ineligible or nonconsenting app installers (*nonparticipants*) could receive the study intervention, which was defined as the degree of app usage. Eligible nonparticipants were not asked to explain why they declined to participate in the study. Outcomes included app usage (as defined by number of times participants listened to guided imagery audio files, number of times participants answered the daily questions, and so on) for all users and smoking status of participants at follow-up. Incentives of US $25 were provided to participants

upon completing each 30- and 90-day assessment. Users could receive app-based awards if they met their goals for a week, but they did not receive monetary compensation.

Basic profile characteristics (eg, demographics and tobacco use) were obtained for all app installers at registration, and additional baseline variables were assessed for participants. Participant outcomes were evaluated at 30 and 90 days post enrollment. A full description of the assessment questionnaire was provided in Gordon et al [9]. App usage was tracked continuously during the trial (April 1, 2015 to October 15, 2015) via app-based analytics (participants only) and Google Analytics (all installers). Participants were asked to set a quit date during app setup but were able to defer setting the date and return to the quit date tool to set a date. Quit dates were recorded in the tracking system, and awards were given based on the quit date. The system's motivational push messages were tailored to a user's stage in the quitting process. The app delivered daily prompts via push notifications to participants to report their smoking status ("Overall, please rate your cravings today" and "Did you smoke today, even a puff?") and reminders to listen to guided imagery audio files at user-specified times. Four 5-min guided imagery audio files focused on smoking cessation, eating well, increasing physical activity, and maintaining general well-being were delivered consecutively over 4 weeks. A complete description of program content, including prompt texts, is provided in a separate paper [9].

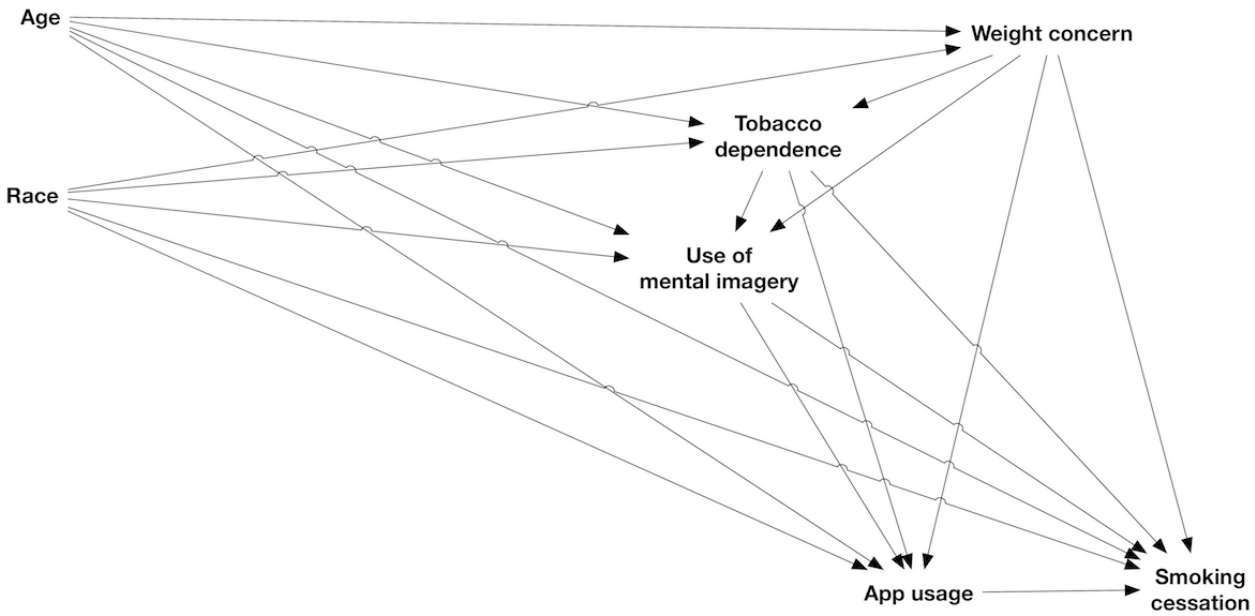
The app was identical for participants and nonparticipants, and both groups were prompted to access app features and complete questions daily; however, only participants were asked to complete the baseline and 30- and 90-day surveys. Participants were required to use the app for at least 30 days during study participation. Staff attempted to contact participants who did not use the app for 2 consecutive weeks, and participants who opted to remain and subsequently used the app were retained. If participants did not use the app for 2 consecutive weeks (14 days) during the first 30 days and the staff was not able to contact them, they were dropped for inactivity. Participants who completed the study were considered *full participants*, whereas those who dropped out or were withdrawn were considered *partial participants*. Data from partial participants were not retained or included in the analysis, though baseline variables were previously reported to not differ significantly between partial and full participants [9].

Study Outcomes and Covariates

Outcomes assessed by this analysis consisted of app usage and smoking cessation indicators. We hypothesized an a priori conceptual relationship model for exposure, outcome, and putative confounder variables (Figure 1). Age and race are commonly considered likely confounders of associations in epidemiological studies. Weight concern was considered likely to influence tobacco dependence [7], whereas both weight concern and tobacco dependence were considered probable influences on motivation to use mental imagery or the app. Guided by this model, separate analyses were used to examine usage as an outcome of demographic and behavioral variables and as an exposure associated with smoking cessation outcomes.

Google Analytics data represent aggregate counts (for participants and nonparticipants separately) of times opening the app, updating quit dates, and starting or completing guided imagery audio files. Outcomes were measured as unique screen views per app session over the entire study period. Audio file data were combined to yield aggregate measures across all audio files. The app collected data in a separate database on audio files completed and days answering questions for each participant.

Figure 1. Directed acyclic graph showing hypothesized relationships among principal exposure, confounder and outcome variables (see text for details of which variables filled these roles in each analysis). App usage and smoking cessation each consist of several measures and are displayed here in a simplified form.



Completion rates for daily individual questions were recorded; however, no monetary reward or compensation was offered for completing the daily assessments. Smoking data were collected at 30 and 90 days and consisted of 7-, 30-, and 90-day self-reported smoking abstinence.

Characteristics collected during app registration for all installers included gender, smoking status in previous 30 days (yes or no), desire to set a quit date (yes or no), number of cigarettes smoked on a typical day, amount paid for a pack of cigarettes, and source of referral to SMSF (from a coded list). Variables with which to address associations between participant characteristics and app usage were sourced from the baseline questionnaire and included age (in years, as 2015 minus reported birth year), race (dichotomized to white or non-white because of small samples for non-white races), tobacco dependence (time to smoking after waking, four categories from <5 min to >60 min), concern about weight gain (low to high willingness to gain 1-5 lb after quitting smoking, on an integer scale of 1 to 5; modeled as a continuous variable given evident linearity with outcome measures), and use of mental imagery in the previous week (never to greater than 30 min, on an integer scale of 1 to 5).

Statistical Methods

All statistical analyses were performed in R version 3.3.1 (R Core Group) [13], and a type I error rate of 0.05 was prespecified for all tests of significance, which were two-sided. Crude differences in baseline characteristics between participants and nonparticipants were evaluated using two-sample *t* tests for continuous variables and Fisher exact tests for categorical variables. Exact Poisson tests were used to assess group-level differences in mean usage rates over the full study period using aggregate Google Analytics data on participation and usage.

Usage variables from app-based analytics displayed right-skewed distributions, supporting the use of negative binomial or Poisson regression models to estimate associations of app usage (days answering questions and number of audio files completed) with participant baseline characteristics. Negative binomial models were compared with equivalent Poisson models using likelihood ratio tests (LRTs) to assess improvement in model fit with the former models. Individual models estimated associations of each outcome variable with age, race, tobacco dependence, weight concern, and use of mental imagery. Adjusted models were specified from putative confounding relationships (Figure 1). Models for tobacco dependence included weight concern as a covariate, and models for mental imagery included weight concern and tobacco dependence. All models included race and age as possible confounders. Age was treated as a continuous variable with either linear or nonlinear terms (three-knot restricted cubic splines) selected for each model via LRTs. Observations with Cook *d* above 4/n (where n is the sample size for the analysis) within a given adjusted model were removed and the model rerun to assess the sensitivity of results to these potentially influential observations.

Associations between app usage and smoking cessation among individual full participants were assessed using separate logistic regression models for each self-reported smoking abstinence outcome. Primary exposure variables were modeled separately and consisted of days answering questions and audio files completed. Usage data for the first 30 days of participation were summed for each individual. Age (linear), race, weight concern, mental imagery, and tobacco dependence were included as potential confounders. Assumptions of linearity in log-odds were assessed, and log transformation and nonlinear modeling of exposure variables were evaluated for improvement of linearity and model fit. Discriminatory abilities of final models were evaluated using calculated *C* indices. Identification of influential points and handling of missing data were as described previously for negative binomial models.

Results

Participant Recruitment and Characteristics

Recruitment of participants into the feasibility trial is illustrated in Figure 2. Of the total of 289 individuals who installed the SMSF app during the recruitment period, 251 completed the registration profile. Of these, 86 individuals chose not to enroll, and 14 were dropped because of exclusion criteria. A total of 151 eligible individuals consented to study participation and were enrolled, of whom 15 requested to be withdrawn from the study, and 63 were dropped because of inactivity, yielding 78 partial participants and 73 full participants. The median time until dropout because of either withdrawal-request or inactivity was 33 days, with a median time until withdrawal-request of 23.5 days (range: 15-48 days; n=15) and median time until dropout because of inactivity of 34 days (range: 4-82 days; n=63).

Baseline characteristics of study participants and nonparticipants are summarized in Table 1. Participants had significantly higher likelihood of female gender and of smoking in the 30 days before enrollment than nonparticipants. Whereas a larger proportion of participants than nonparticipants set a quit date at baseline, and participants smoked a higher mean number of cigarettes per day and paid a higher average price for cigarettes, these differences were not statistically significant. Participants were more likely to have been referred to SMSF by Google or the Google Play Store than nonparticipants, but other sources of referral did not differ significantly.

Figure 2. Flow of participants through the See Me Smoke-Free (SMSF) feasibility trial.

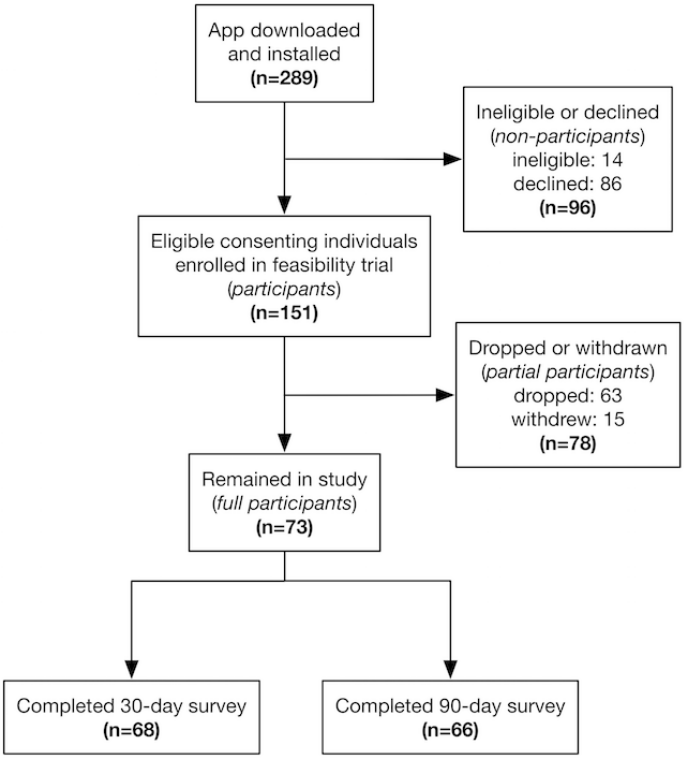


Table 1. Baseline characteristics from registration profile responses for all users who completed a profile; study participants and nonparticipants. Test results are for null hypotheses of no mean difference between participants and nonparticipants.

Profile characteristics	All users (N=247)	Participants (n=151)	Nonparticipants (n=96)	*P* value
	Mean (SD[a]) or n (%)	Mean (SD) or n (%)	Mean (SD) or n (%)	
Gender: female, n (%)	243 (98)	151 (100)	92 (96)	.02[b]
Smoked in previous 30 days, n (%)	237 (96)	151 (100)	86 (90)	<.001[b]
Set quit date at baseline, n (%)	198 (80)	127 (84)	71 (74)	.07[b]
Cigarettes smoked on a typical day, mean (SD)	16.1 (10.6)	16.6 (11.9)	15.2 (8.1)	.27[c]
Amount paid for a pack of cigarettes in US $, mean (SD)	6.61 (2.80)	6.70 (2.77)	6.47 (2.86)	.54[c]
Referred to SMSF [d] by [e], n (%)				
Facebook	29 (12)	19 (13)	10 (10)	.69[b]
Google or Google Play Store	78 (32)	56 (37)	22 (23)	.02[b]
Friend or family	26 (11)	14 (9)	12 (13)	.52[b]
Other	135 (55)	79 (52)	56 (58)	.36[b]

[a]SD: standard deviation.

[b]Fisher exact test.

[c]*t* test (2-tailed).

[d]SMSF: See Me Smoke-Free.

[e]Totals may exceed 100%, as users could provide more than one response.

Age and race data were collected only for participants, and one partial participant did not complete the baseline survey; all other baseline variables of interest were complete, and full and partial participants reported similar baseline characteristics. Age was similar among full and partial participants (full: mean=39.1 years, standard deviation [SD]=13.1 years; partial: mean=36.9 years, SD=11.4 years; *t* test: *P*=.29). Racial composition was also similar among full and partial participants (white: 73% [53/73] and 74% [57/77], respectively; African American: 16% [12/73], 12% [9/77]; Asian: 1% [1/73], 3% [2/77]; multiracial:

4% [3/73], 4% [3/77]; Native American: 0% [0/73], 1% [1/77]; other: 6% [4/73], 7% [5/77]; Fisher exact test: *P*=.94). Full and partial participants had moderate and similar concerns about weight gain (full: mean=3.1, SD=1.1; partial: mean=2.9, SD=1.2; on a scale from 1 to 5; *t* test: *P*=.49) and reported high smoking dependence, with 52 (71%, 52/73) full and 63 (82%, 63/77) partial participants reporting that they smoke within 30 min of waking (*t* test of reported dependence on continuous scale: *P*=.47). Most participants reported using mental imagery only infrequently at baseline, with 42 (58%, 42/43) full and 38 (49%, 38/78) partial participants reporting no use of mental imagery during the previous week (*t* test of reported mental imagery use on continuous scale: *P*=.76). Full and partial participants set quit dates at baseline at similar rates (Fisher exact test: *P*=.83) and reported similar numbers of cigarettes smoked at baseline (*t* test: *P*=.54) and amounts paid for cigarettes (*t* test: *P*=.10).

Participation and Usage

On the basis of group-level usage data from Google Analytics, participants and nonparticipants opened the app and updated quit dates at the same average rate (Table 2), but participants started and completed audio files at significantly higher rates. Relative usage patterns of full versus partial participants could not be distinguished using the Google Analytics data. An in-depth comparison of daily assessment response rates between full versus partial participants was beyond the scope of this analysis; however, a total of 123 participants completed the daily assessment at least once.

Among these respondents (n=123), the daily assessment was completed on an average of 21.2 days (SD=26.6; range: 1-118). In response to the prompt “overall, please rate your cravings today,” users selected “none” on an average of 21 days (SD=25.2; range: 1-87), they selected “few” on an average of 9.2 days (SD=11.1; range: 1-63), they selected “several” on an average of 5.9 days (SD=9.6; range: 1-70), they selected “many” on an average of 3.9 days (SD=4.07, range: 1-21), and they selected “very many” on an average of 4.9 days (SD=9.2, range: 1-48). In response to the prompt “Did you smoke today, even a puff?” users selected “yes” on an average of 18 days (SD=23.7; range: 1-115), and they selected “no” on an average of 8.8 days (SD=13; range: 1-61).

Participant Characteristics and Usage

Full participants listened to audio files on an average of 30.2 times (SD=29.4; range: 0-109) and answered questions on an average of 33.4 days (SD=29.1; range: 0-109). The use of negative binomial regression was supported by the highly right-skewed distributions of both usage variables, the significantly improved fit relative to Poisson models, and unremarkable residual plots. Usage outcomes were highly positively correlated (Pearson correlation ρ=.77). Whereas examined baseline characteristics were not significantly associated with extent of app usage (Table 3), results suggested higher relative usage with increasing age and non-white race and lower usage with increasing weight concern. Increased smoking dependence and use of mental imagery were not associated with consistent trends in app usage. Results did not differ meaningfully between unadjusted and adjusted models. Reanalysis after removal of influential observations reversed the association between race and audio completion and suggested a stronger relationship between increased smoking dependence and daily questions answered. No other estimates were highly sensitive to influential observations.

Table 2. Comparison of app usage outcomes from Google Analytics for participants and nonparticipants. Total: Aggregate count. Mean: Average per user. RR: rate ratio (exact Poisson test), with test for equal rates. Counts of audio files started and completed include both repeated usages of the same files and usages of different files by an individual.

Usage outcome	Participants (n=151)	Nonparticipants (n=96)	RR[a] (95% CI)	*P* value
	Total (mean)	Total (mean)		
Opening the app	2912 (19.3)	1895 (19.7)	0.98 (0.92-1.04)	.43
Updated a quit date	29 (0.19)	18 (0.19)	1.02 (0.55-1.96)	>.99
Audio files started	2725 (18.0)	1625 (16.9)	1.07 (1.00-1.13)	.04
Audio files completed	2288 (15.2)	1315 (13.7)	1.11 (1.03-1.18)	.003

[a]RR: rate ratio.

Table 3. Associations between baseline characteristics and usage outcomes. App-based analytics for full trial participants (n=73) over the study period, as incidence rate ratios (IRR); results of sensitivity analysis, where influential observations are removed, are included.

Characteristic	Daily questions answered, IRR[a] (95% CI)			Audio files completed, IRR (95% CI)		
	Unadjusted	Adjusted[b]	Influential observations removed[c]	Unadjusted	Adjusted	Influential observations removed
Age (per 10 years)	1.05 (0.87-1.26) *P*=.64	1.07 (0.89-1.30) *P*=.46	1.07 (0.89-1.30) *P*=.46, n=73	1.15 (0.96-1.38) *P*=.13	1.18 (0.98-1.42) *P*=.07	1.26 (1.05-1.51) *P*=.01, n=70
Race						
White	Ref	Ref	Ref	Ref	Ref	Ref
Other	1.16 (0.69-2.06) *P*=.59	1.25 (0.71-2.29) *P*=.43	1.25 (0.71-2.29) *P*=.43, n=73	1.02 (0.61-1.77) *P*=.95	1.23 (0.71-2.20) *P*=.44	0.86 (0.50-1.52) *P*=.44, n=70
Time to smoke after waking	*P*=.57	*P*=.72	*P*=.39	*P*=.18	*P*=.67	*P*=.17
<5 min	0.90 (0.31-2.16)	0.92 (0.31-2.33)	1.53 (0.48-4.04) n=71	0.69 (0.26-1.60)	0.81 (0.30-1.96)	0.90 (0.31-2.22) n=68
5-30 min	1.29 (0.46-3.02)	1.28 (0.46-2.99)	2.06 (0.67-5.17) n=71	1.26 (0.48-2.85)	1.19 (0.45-2.69)	1.62 (0.58-3.83) n=68
31-60 min	0.90 (0.30-2.32)	0.90 (0.30-2.37)	1.26 (0.38-3.50) n=71	0.82 (0.29-2.04)	0.89 (0.32-2.20)	0.92 (0.31-2.44) n=68
>60 min	Ref	Ref	Ref	Ref	Ref	Ref
Concern about weight gain[d]	0.89 (0.71-1.12) *P*=.32	0.91 (0.72-1.15) *P*=.43	0.88 (0.69-1.11) *P*=.25, n=68	0.83 (0.66-1.05) *P*=.09	0.85 (0.68-1.06) *P*=.13	0.78 (0.62-0.97) *P*=.02, n=69
Mental imagery use[e]	0.93 (0.76-1.16) *P*=.47	0.93 (0.76-1.17) *P*=.50	0.87 (0.69-1.11) *P*=.19, n=71	0.98 (0.81-1.22) *P*=.85	1.00 (0.83-1.23) *P*=.78	0.84 (0.69-1.05) *P*=.09, n=67

[a]IRR: incidence rate ratios.

[b]Model adjusted for age and race.

[c]n: number of observations remaining after exclusion of influential observations.

[d]1=low to 5=high.

[e]1=Never to 5=More than 30 min.

App Usage and Smoking Cessation

Among 68 full participants who completed the 30-day survey, 7- and 30-day self-reported smoking abstinence was 37% (25/68) and 21% (14/68), respectively. Among 66 full participants who completed the 90-day survey, 7-, 30-, and 90-day self-reported smoking abstinence was 47% (31/66), 32% (21/66), and 15% (10/66), respectively. Cessation outcomes at the 30-day survey were highly positively correlated (Pearson correlation ρ=.77), but cessation outcomes at the 90-day survey and between 30- and 90-day surveys were less highly correlated (ρ=.43-.73). As our study design (within subjects) was focused on looking only at participants who actually used the app, usage data are not available for partial participants.

The data suggested that app usage may have a nonlinear relationship with smoking cessation, but model comparisons supported use of untransformed usage variables in all final analyses. Among full participants, odds ratios for most associations between app usage and smoking cessation were greater than one (Table 4), and some were statistically significant, suggesting a positive association between app usage and smoking cessation. Missing cessation outcomes were not imputed as missingness was low among full participants (7% [5/73] and 10% [7/73] for 30- and 90-day survey outcomes, respectively; partial participants lacked outcome data by definition and were not considered missing for this purpose). Adjusted models had moderate discriminatory ability, with *C* indices ranging from 0.68 to 0.78 (results not shown). Results were similar between unadjusted and adjusted models, though adjusted models yielded larger effect sizes. With the exception of 7-day smoking cessation at the 30-day survey, reanalysis after exclusion of influential observations also suggested stronger associations between app usage and smoking cessation than the primary analysis.

Table 4. Associations between app usage during the first 30 days and smoking cessation outcomes from 30- and 90-day surveys for full trial participants. To provide meaningful comparisons, estimates are given for the third versus first quartiles of the main exposures. Influential observations removed: Results of sensitivity analyses (adjusted models).

Self-reported smoking abstinence	Complete audio listens, OR[a] (95% CI)			Days answering questions, OR (95% CI)		
	Unadjusted	Adjusted[b]	Influential observations removed[c]	Unadjusted	Adjusted	Influential observations removed
30-day survey (n=68)						
7 days	1.57 (0.62-3.95) *P*=.34	1.78 (0.61-5.23) *P*=.29	1.51 (0.35-6.48) *P*=.58, n=63	1.28 (0.49-3.31) *P*=.62	1.63 (0.52-5.14) *P*=.40	1.21 (0.26-5.62) *P*=.81, n=63
30 days	1.31 (0.44-3.87) *P*=.63	1.33 (0.37-4.69) *P*=.66	4.35 (0.36-52.08) *P*=.25, n=61	0.79 (0.25-2.52) *P*=.69	0.86 (0.22-3.38) *P*=.83	0.87 (0.10-7.63) *P*=.90, n=61
90-day survey (n=66)						
7 days	2.84 (1.05-7.66) *P*=.04	3.04 (1.07-8.67) *P*=.04	4.19 (1.36-12.97) *P*=.01, n=62	2.54 (0.96-6.67) *P*=.06	3.17 (1.07-9.38) *P*=.04	6.27 (1.62-24.27) *P*=.01, n=62
30 days	2.37 (0.87-6.46) *P*=.09	3.06 (1.00-9.35) *P*<.05	9.12 (1.99-41.79) *P*<.01, n=61	1.79 (0.67-4.79) *P*=.25	2.55 (0.82-7.95) *P*=.11	3.87 (1.09-13.73) *P*=.04, n=64
90 days	1.93 (0.55-6.77) *P*=.30	2.91 (0.63-13.43) *P*=.17	3.61 (0.83-15.80) *P*=.09, n=56	1.03 (0.29-3.70) *P*=.96	1.65 (0.36-7.65) *P*=.52	N/A[d,e]

[a]OR: odds ratio.

[b]Models adjusted for age, race, mental imagery, weight concern, and tobacco dependence.

[c]n: number of observations remaining after exclusion of influential observations.

[d]N/A: not applicable.

[e]Results were unreliable as nearly all (9 of 10) individuals with reported smoking cessation were flagged as influential.

Discussion

Principal Findings

The results of this study supported many hypothesized associations in the SMSF trial. Participants were significantly more likely to report that they had smoked during the 30 days before enrollment and to report female gender, both of which were requirements of study enrollment. These results suggest that the app may have appeal beyond the target group. One aim of this study was to explore how the sample of participants compared with users who may or may not have met the inclusion criteria and how other people who downloaded the app (eg, men and those not ready to quit) used the app. As participants were more likely than nonparticipants to have learned about SMSF from Google or the Google Play Store, it is possible that participants were more highly motivated to quit smoking and to have actively sought out cessation-related tools or products. Results also suggested possible associations between usage and both age and weight concern, suggesting a need for further investigation to understand the relationship between these variables. Age has been observed to be a predictor of usage of smoking cessation websites [14]; however, associations with mobile phone–based cessation programs have been mixed, as noted by Zeng et al [10]. Regarding weight concern, the SMSF app was targeted at women smokers who were concerned about weight gain, which is why it includes diet and physical activity components aimed at reducing weight gain while quitting smoking. Study participants did not open the app more frequently than nonparticipants but did start and complete audio files at a significantly higher rate. Conclusions about the clinical meaningfulness of these differences cannot be made because the same outcome measures for participants and nonparticipants were not collected. A positive dose-response relationship to the number of audio files listened to, and smoking abstinence, has been identified by an earlier analysis of the SMSF program [9]. Most of the analyses suggested that participants who used the app more frequently were also more likely to quit smoking, consistent with the findings of Gordon et al [9], who reported significant increases in 7- and 30-day self-reported smoking abstinence during follow-up along with improvements in certain aspects of physical activity and diet.

The contribution of SMSF to the field of mHealth as an innovative imagery-based, multi-behavioral intervention has been discussed in earlier publications [7-9]. Uncertainty often surrounds the extent to which the use of digital interventions determines desired health outcomes [5]. A review by Donkin et al [15] discussed how the most appropriate metrics of usage may differ between different types of interventions and how in-depth analysis of usage can help understand which metrics are most associated with effectiveness. Additionally, user characteristics can strongly influence the effectiveness of digital interventions [6]. This study sought to determine characteristics

of users in the SMSF feasibility trial, how these characteristics might be related to app usage, and to assess whether there is preliminary evidence that app usage is associated with self-reported smoking abstinence.

Strengths and Limitations

Limitations of the SMSF feasibility study have been discussed by Gordon et al [9]. Potential sources of bias include the lack of a control group and the significant dropout rate. Among those who used the app for the duration of the study and responded to follow-up measures, app use was positively associated with cessation. Retention on this study is comparable with other mHealth studies, which frequently have high attrition rates [16-19]. The high attrition inherent in many mHealth studies is offset by the broad reach and dissemination potential of mHealth intervention programs. Participants were recruited from a self-selected pool of individuals who responded to project news coverage and social media promotions, raising the risk of selection bias. Respondents could have differed from the broader target population of female smokers, for example, by being more highly motivated to quit and more inclined to engage in mHealth interventions. The majority of full participants self-identified as white, possibly limiting generalizability to other populations or indicating a need to target the app to less represented groups [20]. Furthermore, SMSF is available exclusively on Android phones, potentially narrowing the user base by socioeconomic status [15]. Despite these limitations, the SMSF trial had several strengths, including the analysis of multiple independent sources of usage data. Whereas some results were sensitive to a small number of influential participants, an additional strength of the study was the general robustness of most estimates. Important additional factors (variables, interactions, or nonlinear relationships) determining app usage or smoking cessation may remain unidentified, but the large effect sizes estimated for most associations between app usage and cessation outcomes are promising and support more extensive evaluation of SMSF efficacy. We note that as a feasibility study, this trial's primary objective was not to detect intervention effects, so results should be interpreted cautiously.

Conclusions

This study suggests high potential efficacy of the SMSF app, as increased usage was generally associated with higher self-reported smoking abstinence. As a one-arm feasibility trial, a causal relationship between app usage and improved smoking cessation cannot be demonstrated, and the study was not powered to identify significant associations. A planned SMSF controlled trial should provide additional evidence with which to judge the app's efficacy as an intervention tool and afford greater statistical power and may be able to examine additional factors affecting app usage and smoking cessation.

Acknowledgments

This study was funded by a grant from the National Cancer Institutes (R21 CA174639; J. Gordon, PI).

Authors' Contributions

CAS and JKR contributed equally to the conceptualization of the study design and methods, conduct of the analyses, and writing of the manuscript. JA participated in the writing of the manuscript. MLB provided oversight of the statistical analyses and edited the manuscript. JSG provided overall supervision of the study design, methods, and analyses and participated in the writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Abroms LC, Whittaker R, Free C, Mendel VAJ, Schindler-Ruwisch JM. Developing and pretesting a text messaging program for health behavior change: recommended steps. JMIR Mhealth Uhealth 2015 Dec;3(4):e107 [FREE Full text] [doi: 10.2196/mhealth.4917] [Medline: 26690917]
2. Jha P, Peto R. Global effects of smoking, of quitting, and of taxing tobacco. N Engl J Med 2014 Jan 02;370(1):60-68. [doi: 10.1056/NEJMra1308383] [Medline: 24382066]
3. ITU. ICT facts & figures URL: http://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx [accessed 2017-04-12] [WebCite Cache ID 6pgABzNWf]
4. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016(4):CD006611. [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
5. Mattila E, Lappalainen R, Välkkynen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. JMIR Mhealth Uhealth 2016 Jul 28;4(3):e90 [FREE Full text] [doi: 10.2196/mhealth.5241] [Medline: 27468653]
6. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
7. Giacobbi P, Hingle M, Johnson T, Cunningham JK, Armin J, Gordon JS. See me smoke-free: protocol for a research study to develop and test the feasibility of an mhealth app for women to address smoking, diet, and physical activity. JMIR Res Protoc 2016 Jan 21;5(1):e12 [FREE Full text] [doi: 10.2196/resprot.5126] [Medline: 26795257]

8. Armin J, Johnson T, Hingle M, Giacobbi P, Gordon JS. Development of a multi-behavioral mhealth app for women smokers. J Health Commun 2017 Feb;22(2):153-162. [doi: 10.1080/10810730.2016.1256454] [Medline: 28121240]
9. Gordon JS, Armin J, Giacobbi P, Cunningham JK, Johnson T, Abbate K, et al. Development and evaluation of the see me smoke-free multi-behavioral mhealth app for women smokers. Transl Behav Med 2017;7(2):172-184. [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
10. Zeng EY, Vilardaga R, Heffner JL, Mull KE, Bricker JB. Predictors of utilization of a novel smoking cessation smartphone app. Telemed J E Health 2015;21(12):998-1004. [doi: 10.1089/tmj.2014.0232] [Medline: 26171733]
11. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan;41(1):68-73. [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
12. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [Medline: 27777223]
13. R-project. 2016. R: a language and environment for statistical computing URL: https://www.R-project.org/ [accessed 2017-09-06] [WebCite Cache ID 6tHi2DiF6]
14. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in internet-based interventions. Patient Prefer Adherence 2008 Feb 02;2:57-65 [FREE Full text] [Medline: 19920945]
15. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
16. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an ehealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
17. Pellegrini CA, Verba SD, Otto AD, Helsel DL, Davis KK, Jakicic JM. The comparison of a technology-based system and an in-person behavioral weight loss intervention. Obesity (Silver Spring) 2012 Feb;20(2):356-363 [FREE Full text] [doi: 10.1038/oby.2011.13] [Medline: 21311506]
18. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17(1):e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
19. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2016 Nov 04;170:37-42. [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
20. Rahmati A, Tossell C, Shepard C, Kortum P, Zhong L. Exploring iPhone usage: the influence of socioeconomic differences on smartphone adoption, usage and usability. 2012 Presented at: 14th international conference on Human-computer interaction with mobile devices and services; September 21-24, 2012; San Francisco, CA p. 11-20. [doi: 10.1145/2371574.2371577]

Abbreviations

IRR: incidence rate ratio
LRTs: likelihood ratio tests
mHealth: mobile health
OR: odds ratio
RR: rate ratio
SD: standard deviation
SMSF: See Me Smoke-Free

Edited by C Dias; submitted 21.04.17; peer-reviewed by M Businelle, M allman-Farinelli; comments to author 09.05.17; revised version received 08.08.17; accepted 14.08.17; published 09.10.17.

Please cite as:
Schmidt CA, Romine JK, Bell ML, Armin J, Gordon JS
User Participation and Engagement With the See Me Smoke-Free mHealth App: Prospective Feasibility Trial
JMIR Mhealth Uhealth 2017;5(10):e142
URL: https://mhealth.jmir.org/2017/10/e142/
doi:10.2196/mhealth.7900
PMID:28993302

Original Paper

Feasibility of Gamified Mobile Service Aimed at Physical Activation in Young Men: Population-Based Randomized Controlled Study (MOPO)

Anna-Maiju Leinonen[1,2,3], MSc; Riitta Pyky[1,3,4,5], MSc; Riikka Ahola[1,6], PhD; Maarit Kangas[1,4], PhD; Pekka Siirtola[7], PhD; Tim Luoto[8], MSc; Heidi Enwald[4,9], PhD; Tiina M Ikäheimo[4,10], PhD; Juha Röning[2,7], PhD; Sirkka Keinänen-Kiukaanniemi[4,5,11], MD, PhD; Matti Mäntysaari[12], MD, PhD; Raija Korpelainen[3,4,5], PhD; Timo Jämsä[1,2,4,13], PhD

[1]Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
[2]Infotech Oulu, University of Oulu, Oulu, Finland
[3]Oulu Deaconess Institute, Department of Sports and Exercise Medicine, Oulu, Finland
[4]Medical Research Center, Oulu University Hospital and University of Oulu, Oulu, Finland
[5]Center for Life Course Health Research, University of Oulu, Oulu, Finland
[6]Polar Electro, Kempele, Finland
[7]Faculty of Information Technology and Electrical Engineering, Biomimetics and Intelligent Systems Group, University of Oulu, Oulu, Finland
[8]Department of Cultural Anthropology, Faculty of Humanities, University of Oulu, Oulu, Finland
[9]Department of Information and Communication Studies, Faculty of Humanities, University of Oulu, Oulu, Finland
[10]Center for Environmental and Respiratory Health Research, University of Oulu, Oulu, Finland
[11]Health Center of Oulu, Oulu, Finland
[12]Center for Military Medicine, The Finnish Defence Forces, Helsinki, Finland
[13]Diagnostic Radiology, Oulu University Hospital, Oulu, Finland

Corresponding Author:
Anna-Maiju Leinonen, MSc
Research Unit of Medical Imaging, Physics and Technology
University of Oulu
PO Box 5000
Oulu, 90014 University of
Finland
Phone: 358 29 448 600
Fax: 358 8537 5111
Email: anna.jauho@oulu.fi

Abstract

Background: The majority of young people do not meet the recommendations on physical activity for health. New innovative ways to motivate young people to adopt a physically active lifestyle are needed.

Objective: The study aimed to study the feasibility of an automated, gamified, tailored Web-based mobile service aimed at physical and social activation among young men.

Methods: A population-based sample of 496 young men (mean age 17.8 years [standard deviation 0.6]) participated in a 6-month randomized controlled trial (MOPO study). Participants were randomized to an intervention (n=250) and a control group (n=246). The intervention group was given a wrist-worn physical activity monitor (Polar Active) with physical activity feedback and access to a gamified Web-based mobile service, providing fitness guidelines, tailored health information, advice of youth services, social networking, and feedback on physical activity. Through the trial, the physical activity of the men in the control group was measured continuously with an otherwise similar monitor but providing only the time of day and no feedback. The primary outcome was the feasibility of the service based on log data and questionnaires. Among completers, we also analyzed the change in anthropometry and fitness between baseline and 6 months and the change over time in weekly time spent in moderate to vigorous physical activity.

Results: Mobile service users considered the various functionalities related to physical activity important. However, compliance of the service was limited, with 161 (64.4%, 161/250) participants visiting the service, 118 (47.2%, 118/250) logging in more than once, and 41 (16.4%, 41/250) more than 5 times. Baseline sedentary time was higher in those who uploaded physical activity data until the end of the trial (*P*=.02). A total of 187 (74.8%, 187/250) participants in the intervention and 167 (67.9%, 167/246) in the control group participated in the final measurements. There were no differences in the change in anthropometry and fitness from baseline between the groups, whereas waist circumference was reduced in the most inactive men within the intervention group (*P*=.01). Among completers with valid physical activity data (n=167), there was a borderline difference in the change in mean daily time spent in moderate to vigorous physical activity between the groups (11.9 min vs −9.1 min, *P*=.055, linear mixed model). Within the intervention group (n=87), baseline vigorous physical activity was inversely associated with change in moderate to vigorous physical activity during the trial (*R*=−.382, *P*=.01).

Conclusions: The various functionalities related to physical activity of the gamified tailored mobile service were considered important. However, the compliance was limited. Within the current setup, the mobile service had no effect on anthropometry or fitness, except reduced waist circumference in the most inactive men. Among completers with valid physical activity data, the trial had a borderline positive effect on moderate to vigorous physical activity. Further development is needed to improve the feasibility and adherence of an integrated multifunctional service.

Trial registration: Clinicaltrials.gov NCT01376986; http://clinicaltrials.gov/ct2/show/NCT01376986 (Archived by WebCite at http://www.webcitation.org/6tjdmIroA)

(JMIR Mhealth Uhealth 2017;5(10):e146) doi:10.2196/mhealth.6675

KEYWORDS

accelerometry; adolescent; behavior change; health; Internet; self-monitoring; wearable

Introduction

The positive effect of physical activity on health [1-3], fitness [4,5], and other lifestyle factors, such as smoking and irregular eating [6], is undeniable. To achieve the health benefits of physical activity, young people under the age of 18 years should accumulate at least 60 min of moderate to vigorous physical activity daily [2]. However, the majority of young people do not meet the global recommendations of physical activity beneficial for their health [7]. Physical activity declines during adolescence even more among boys than girls [8]. Thus, encouraging physical activity among young men is of great importance, and new innovative solutions to motivate young men to adopt a physically active lifestyle are needed [8,9].

Traditional interventions, including face-to-face meetings, consume time and money and are also associated with geographic restrictions [10,11]. On the other hand, interventions delivered through mass media are not able to provide individualized feedback, which has been shown to enhance the effectiveness of physical activity-related interventions [12]. The constantly increasing availability of the Internet and mobile apps offer a new viable way to reach young people easily, even with tailored feedback [13,14]. In addition to tailored and real-time feedback, physical activity profiles, goal setting, social support networking, and online expert consultation have been found to be effective ways to enhance physical activity with smartphone technology [15].

Several previous studies investigating the effect of an Internet-based service or a mobile service for promoting physical activity have shown improved health-related behavior among study participants [16-21]. However, the use of questionnaires to evaluate change in physical activity and the lack of a comparable control group have complicated the investigation of the effect of interventions [15,22]. In addition, intervention studies targeted at young people, especially interventions targeted at boys, are still scarce [12,23].

Nowadays, technology such as wearables enables integrating measured sensor data (eg, physical activity data) directly as parts of mobile services, which further allows delivering automated and tailored feedback messages to the user based on the personal measures. Tailoring of health communication is a means to increase effectiveness of health information by providing more user-centered information [24]. New technology also allows integrating game mechanics to the nongame contexts in a service.

Mobile phone games can be feasible for adolescents to use for promoting physical activity [25]. Gamification means using game design techniques, game principles, and mechanics, such as badges, points, levels, and leaderboards, to improve user engagement, learning, behavior change, and reaching goals [26]. Gamification is increasingly being used for fitness and health-related services to improve user experience and engagement [27], and there are number of health and fitness apps available in the app stores containing at least some components of gamification. Recent reviews have provided an overview of mobile gaming apps to promote daily life physical activity and a demonstration of their acceptability and feasibility among the users [28] and have identified and confirmed the effectiveness of persuasive features in physical activity studies [29]. However, clinical effectiveness and the added value of gaming in changing daily activity behavior have not yet been established. Furthermore, integration of elements of behavioral theory is lacking, which can potentially impact the efficacy of gamification apps to change behavior [30].

In our 3-month MOPO pilot study, we demonstrated that the use of a wearable physical activity monitor providing activity feedback had a short-term positive effect on physical activity behavior among young men [31]. To meet the need for new

methods to motivate young people to promote their physical activity, this 6-month trial (ClinicalTrials.gov NCT01376986) evaluated the feasibility of a fully automated, gamified, tailored Web-based mobile service, including physical activity monitoring, in young men. Additionally, we analyzed the effects of the service on anthropometry, fitness, and physical activity among completers. We hypothesized that the mobile service together with continuous physical activity monitoring and feedback is feasible for young men and that the service has a positive effect on anthropometry, fitness, and physical activity.

Methods

Study Design and Participants

This 6-month, population-based, parallel randomized controlled trial (MOPO study) was conducted in the city of Oulu, Northern Finland (the number of inhabitants approximately 199,000). The recruitment of the study participants was carried out during the annual military call-ups in September 2013. The military service or civic duty is compulsory for all male citizens in Finland. Finnish Defence Forces organize conscription every year, and all boys turning 18 years participate. During the call-ups in autumn 2013, all conscription-aged men (n=1265) were invited to anthropometry and fitness measurements (Figure 1). All those 825 men (65.21%, 825/1265) who went through the measurements were asked to participate in the 6-month trial. Finally, a total of 496 young men (mean 17.8 years [standard deviation 0.6]) agreed to participate and were randomly allocated (allocation ratio 1:1) to either an intervention group (n=250) or a control group (n=246). Blinded randomization was performed by an assistant who was neither involved in the trial nor in the data collection and analysis. Randomization was conducted based on a list of computer-generated random numbers in blocks of 10. Each participant received sequentially the next random assignment in the list. The reasons for not participating in the trial (n=329) were their lack of interest or laziness (31%); they would not use the wrist-worn clock because they disliked its outlook or already had a wrist watch (23%); they did not feel a need for this type of service; they were already taking care of themselves (9%); and other reasons (17%). In addition, 20% did not give any reason for declining.

The study protocol [32] has been registered to the clinical trials register (NCT01376986, ClinicalTrials.gov). The study participants were provided written and oral information about the procedures of the study, and a written consent was obtained. The study was conducted in accordance to the Declaration of Helsinki. The study was approved by the Local Ethics Committee.

Intervention

The men in the intervention group had access to an automated, gamified, tailored Web-based mobile physical and social activation service during the trial. The detailed description of the intervention has been published elsewhere [33].

During the baseline assessments, all participants received a wrist-worn physical activity monitor (Polar Active, Polar Electro). In the intervention group, a personal account for a gamified Web-based mobile service, MOPOrtal (Figure 2), was generated. If necessary, the participants were also provided a mobile phone (n=19) for the duration of the study to be able to use the service. During the first (baseline) week of the trial, the Polar Active monitor was not providing any feedback to the user, and access to the MOPOrtal service was also blocked. After the baseline week, the intervention group was sent a text message (short message service, SMS) instructing how to unlock the monitor screen and log in to MOPOrtal. At the final measurements, the monitors were checked to ascertain whether unlocking had been performed.

Through the trial, the physical activity of the men in the control group was measured continuously with a blinded Polar Active monitor providing only the time of day but no feedback. Otherwise they continued their normal life. The control group had no access to the MOPOrtal service.

All study participants filled in a health and lifestyle questionnaire and went through anthropometry and fitness measurements at baseline in September 2013 and at the end of the trial in March 2014. Those two meetings at baseline and at 6 months were the only face-to-face meetings during the 6-month trial.

Physical Activity Monitor

Polar Active is a wrist-worn watch-like monitor displaying by default the accumulated daily moderate to vigorous physical activity time and achievement of daily activity target (60 min in this study) as a bar. The time spent on different physical activity levels, steps, and calories for each day are also available for the user through the monitor. The monitor with a 21-day memory is waterproof and includes a uniaxial accelerometer. Polar Active calculates the acceleration signals to metabolic equivalents (MET) with the epoch length of 30 s using sex, age, weight, and height as input. In addition, Polar Active provides time spent in five activity levels using the following thresholds: 1≤ MET<2 (sedentary behaviors), 2≤MET<3.5 (light physical activity), 3.5≤MET<5 (moderate physical activity), 5≤MET<8 (vigorous physical activity), and ≥8 MET (very vigorous physical activity). While assessing energy expenditure, a high correlation has been found between Polar Active and the doubly labeled water technique (R=.86), as well as between Polar Active prototype and indirect calorimetry (R=.987) [34,35].

The participants in both groups were advised to wear the device on the nondominant wrist at least for all waking hours and to upload their personal activity data to the research database through Polar FlowLink (Polar Electro) at least every 3 weeks. As a reminder to upload the physical activity data, both groups received a text message every 3 weeks. Two movie tickets were raffled once a month as incentives among those participants who uploaded physical activity data.

At least 3 valid days (≥500 min of data) out of 7 were required to be included in the analysis for each week [36]. Mean daily time was calculated for each week for time spent in moderate (3.5-5 MET) and vigorous physical activity (>5 MET) for both groups starting from the next day when the monitor was given. Moderate to vigorous physical activity was defined as the sum of moderate and vigorous physical activity.

Figure 1. The flow diagram of participants in the 6-month randomized controlled MOPO study.

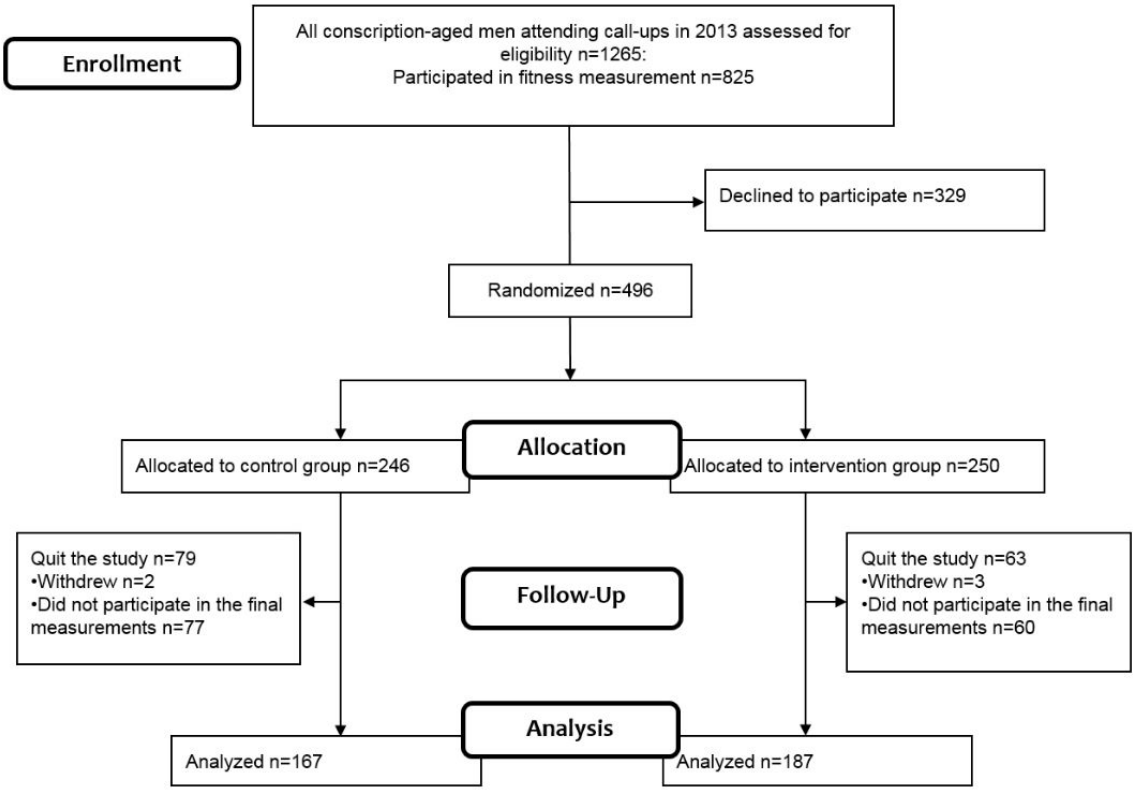


Figure 2. The MOPOrtal service and the Clans of Oulu game.

Gamified Mobile Service

The novel gamified mobile service, MOPOrtal (Figure 2), for promoting physical activity and health was set up in the multidisciplinary MOPO study together with the city of Oulu and enterprises of related expertise, including a game studio (LudoCraft Ltd). Additionally, men in the age group of 16 to 20 years from local school classes, voluntary courses, and youth workshops for unemployed young men were involved in designing and testing the service. The underlying idea of the service design was to promote physical and social activity through game-based persuasion, for example, by physically moving within the districts of the city, players could earn points and claim areas for their clan in-game [37]. A sense of affinity was intended to be achieved via a multiplayer cooperation element, as the player belonged to any of the five clans contesting in the game.

The service was running on a Web browser but optimized for mobile use (HTML5), which enabled participants in the intervention group to use it either on a computer or on a mobile device. The activity data, measured using Polar Active, was utilized from the research database to tailor the feedback and information provided by MOPOrtal for each user. The users had also an opportunity to enter daily activity data manually to the service.

The gamified MOPOrtal service has been described in more detail elsewhere [33]. In short, it included (1) automated tailored health information, exercise, and physical activity instructions based on the stage of exercise behavior change; (2) regular automated change check concerning the stage of exercise behavior change; (3) feedback on physical activity and sitting time; (4) *Clans of Oulu* conquering game based on a map and global positioning system (GPS; see below); (5) social networking; and (6) interface to communal youth services and a Web-based helpdesk for technical issues.

Textual and graphical physical activity feedback provided by the service was based on weekly and daily activity metrics (Figure 3). The weekly feedback was given when activity data of at least 3 days out of 7 was available, otherwise the service reminded the user to upload physical activity data from the monitor. The feedback was based on a comparison of user's activity data to the global recommendation on physical activity for health (60 min/day). Additionally, self-referenced comparison from the previous week was used. Furthermore, the users received positive feedback if their physical activity level was better compared with the weekly average of the whole intervention group (peer-referenced comparison). The overall feedback of daily activity was provided showing a thumb either up, sideways, or down depending on a fulfillment of the global physical activity recommendation and whether the day included over 2 hours of sedentary (sitting) periods or not. In addition, the user was able to see accumulated minutes in different physical activity levels for every 2-hour periods for each day.

The game (*Clans of Oulu* [38]) was based on the location of a person tracked using GPS and played in groups in five different clans using a mobile phone. The basic idea was that by moving physically within the districts of the city of Oulu, players could conquer areas for their own clan. In contrast to traditional games for promoting physical activity, many kinds of activities were rewarded by delivering more points to conquer new areas. For example, uploading of the personal physical activity measures to the research database, fulfillment of daily physical activity recommendation, as well as increment in weekly physical activity and decrement in weekly sedentary time of the player were rewarded. Additionally, by reading facts and health information delivered by the service and inviting friends to join the game, the player received new points to the game. *Clans of Oulu* included, for example, the following game design elements: competition (personal ranks and team ranks), conflict (tasks to be solved and combats with another team), collaboration (working together to reach goals to conquer areas), strategy (points earned based on activity), chance (random new tasks), esthetics (visual appearance for each clan representing different youth cultures), theme (clan game, youth cultures, and conquering), resources (points as resource for concurring areas), time (outdating of points), and scoring or rewards (points for physical activity, completing tasks, and lottery). A more detailed description of the game, its development, and user experiences can be found elsewhere [37].

Gamification was used in MOPOrtal throughout the service, for example, using similar visual appearance as in the *Clans of Oulu* game, and the main actions which occurred in the game (eg, an area occupation from the own clan) were displayed on the service without the need to log in to the game itself.

Feasibility of the Service

The user-specific information concerning the log-ins to the service and the use of the different service sections were recorded to the research database [39]. In addition, during the final measurements, the participants filled in a questionnaire with questions related to the service and its different components.

The Transtheoretical Model of Behavior Change

The health information and feedback delivered by the service was tailored based on the transtheoretical model of behavior change (TTM) [40]. The model *is one of the most popular behavior change models utilized in tailored health interventions [41]*. Originally, it was developed for smoking cessation, but over the last decades, it has also been used in the background of intervention studies aimed to change exercise and physical activity behavior [42,43]. The model includes five different stages for physical activity adaptation and maintenance: precontemplation, contemplation, preparation, action, and maintenance [40].

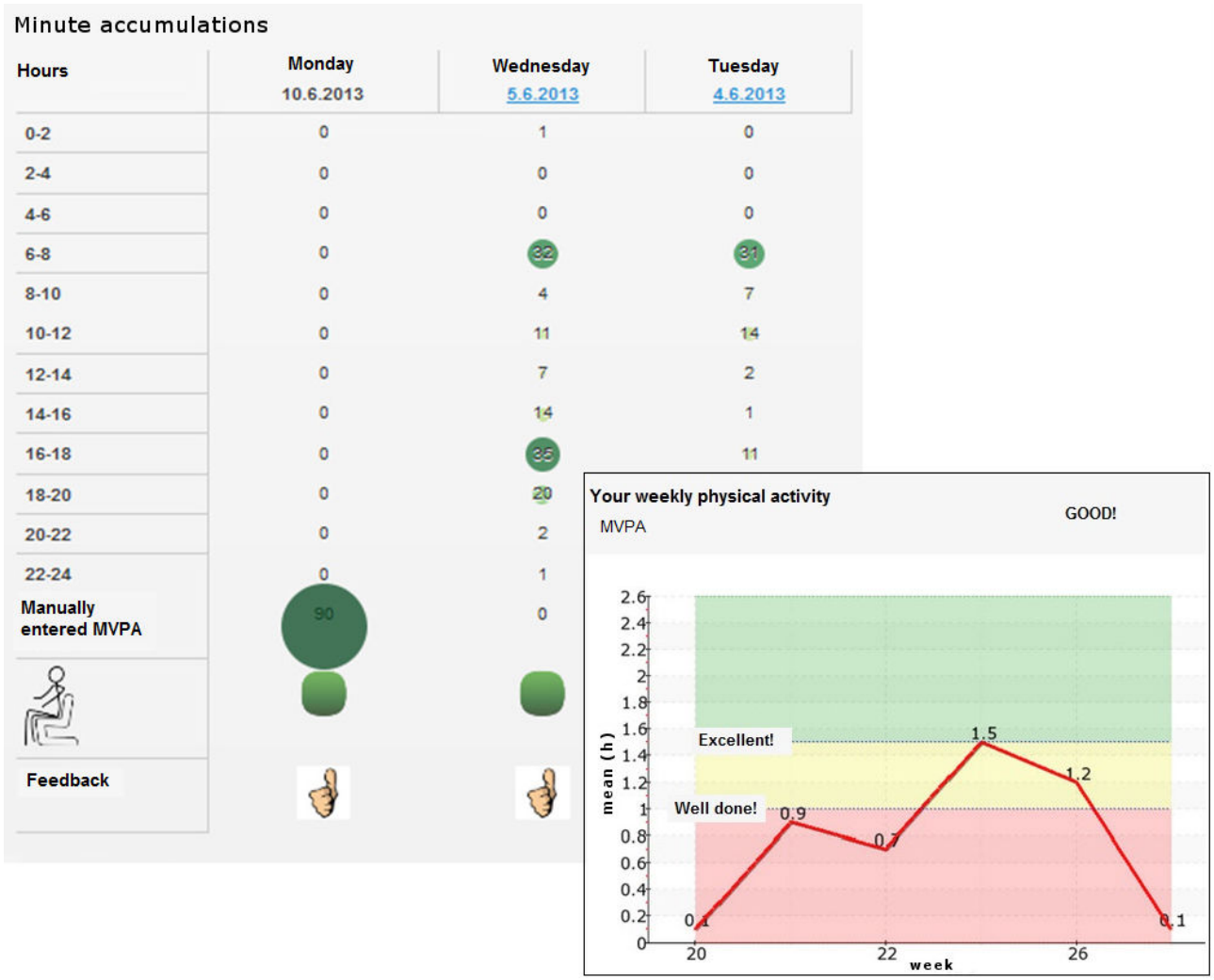
Other core constructs of the model are processes of change, decisional balance, and self-efficacy. As individuals proceed to further stages, their confidence in their ability to sustain a target behavior in various situations (self-efficacy) increases, and the advantages of behavior change outweigh the disadvantages leading to decisional balance. The processes of change represent the type of activities that are initiated or experienced by an individual in their attempt to modify affect, behavior, cognitions, or relationships. [40]. Nine processes of change, which include consciousness-raising, dramatic relief, self-reevaluation, and social liberation and the behavioral processes of

counterconditioning, helping relationships, reinforcement management, self-liberation, and stimulus control, have received the most empirical support in the context of exercise behavior change [44].

The health information (eg, “Did you know that exercising in nature reduces stress more than exercising in urban environment?”) and individually tailored feedback (eg, “Great! You have sufficient physical activity for your health. If you increase your activity further, your fitness will improve.”) messages delivered automatically by the service were tailored to match the processes of change theorized as most appropriate at each stage [44]. Some modifications were made to the design guide derived from the one designed by Nigg et al. [44]. Furthermore, based on research on feedback perceptions of young men in different stages of exercise behavior change [45,46], message tactics that include self- or peer-referenced comparison, namely normative and ipsative messages, were used only in the more advanced stages (action and maintenance).

In this study, the stage of change was assessed during the first visit to the mobile service based on a modified scale from Cardinal (1995) [47]. The respondents were instructed to choose an alternative that best described their regular exercise behavior and intentions to exercise. Regular exercise was defined according to the Finnish national recommendations for those in the age group of 13 to 18 years as at least 1.5 hours of daily physical activity, of which half should be performed at a vigorous intensity [48]. The answer options were (1) I exercise on a regular basis and have been doing so for longer than 6 months (maintenance), (2) I exercise on a regular basis but I have only begun doing so within the past 6 months (action), (3) I do not exercise, but I have been thinking about starting to exercise within the next month (preparation), (4) I do not exercise, but I have been thinking about starting to exercise within the next 6 months (contemplation), and (5) I do not exercise and do not plan to start exercising in the next 6 months (precontemplation). Additionally, the participants were automatically asked to update their stage of change every 2 months.

Figure 3. Examples of graphical physical activity and sitting feedback provided by the MOPOrtal service. The line plot represents the weekly feedback (red area—user’s mean average daily moderate to vigorous physical activity (MVPA) is below the global recommendation (60 min/day); yellow and green area—user’s daily moderate to vigorous physical activity level corresponds to the recommendation). The minute accumulations of moderate to vigorous physical activity are given at every 2-hour periods, with the numbers marked as green indicating high accumulation. The thumb gives feedback indicating whether the daily recommendation was fulfilled or not.



Anthropometry and Fitness

Height and waist circumference (midway between the lowest rib and the iliac crest) were measured with a measuring tape with an accuracy of 0.5 cm. Body composition and weight were assessed by bioelectrical impedance assessment using InBody720 (Biospace Co, Ltd). Bilateral maximal isometric grip strength was measured with a dynamometer while the subject was standing with legs apart and elbow at a 90° angle (SAEHAN Corporation) [49]. The best result of two attempts per hand was recorded. The mean value of both hands was used in the analysis. Polar Fitness Test conducted at rest using FT80 heart rate monitor (Polar Electro) was used to evaluate aerobic fitness. The test assesses maximal oxygen uptake (mL/kg/min) from resting heart rate, heart rate variability, and demographic variables [50]. Polar Fitness Test has been compared with open circuit spirometry for measuring aerobic fitness among adult men with high correlation (.71) and high accuracy (standard error of estimate, SEE=8.5 mL/kg/min) [51]. Additionally, when tested among trained males, the results of Polar Fitness Test were associated highly with laboratory measures (60.2 vs 62.5 mL/kg/min, SEE=7.6 mL/kg/min) [52].

Outcomes

The primary outcomes were the feasibility of the mobile service. The secondary outcomes were changes from baseline in the anthropometry and fitness and the change in weekly time spent in moderate to vigorous physical activity.

Statistical Analysis

The results were analyzed with the Statistical Package for the Social Sciences (SPSS) version 19 (IBM Corp) for Windows software. A *P* value below .05 was considered statistically significant. Anthropometry, fitness, and physical activity variables were tested for normality with the Kolmogorov-Smirnov test. The statistical significance of the differences at baseline in continuous variables between the intervention and control groups, between the study participants and nonparticipants, as well as between those young men who based on the log data visited MOPOrtal service at least once and those who did not visit the service at all during the trial were analyzed using the independent samples *t* test. The within-group changes from baseline in the intervention and control groups were analyzed using the paired samples *t* test.

The difference over time in the change in moderate to vigorous physical activity between the intervention and control groups was analyzed using multiple linear mixed model with full maximum likelihood, compound symmetry, and Bonferroni correction. All available personal weekly averages of mean daily time spent in moderate to vigorous physical activity for both study groups were included in the mixed model analyses.

The Pearson correlation coefficient (*R*) was used to evaluate which variables measured at baseline were significantly associated with the main outcome measure in the intervention group. The association between the usage frequency of the service and the occurred change in moderate to vigorous physical activity time was analyzed using the Spearman rank correlation coefficient (ρ).

Results

Overview

The baseline characteristics of the study participants were similar between the intervention and control groups (Table 1). In addition, the study participants did not differ in anthropometry and fitness from those conscription-aged men who only took part in the fitness measurements (n=329) but not the trial (data not shown).

In total, 187 (74.8%, 187/250) men in the intervention group and 167 (67.9%, 167/246) in the control group completed the study and attended the final measurements after the 6-month trial (Figure 1). From all study participants, 142 (28.6%, 142/496) did not participate in the final measurements and were excluded from the final analysis.

Feasibility of the Mobile Service

On the basis of the log data, 161 men (64.4%, 161/250) in the intervention group visited the MOPOrtal service during the trial, 118 (47.2%, 118/250) logged on the service more than once, and 41 (16.4%, 41/250) more than 5 times. In total, 1044 visits were logged (median: 3, range: 1-202). Use rate decreased during the study, being 400 visits during the first month and 69 during the sixth month. A total of 56 participants used the *Clans of Oulu* game in the service. On the basis of the questionnaire, the most common reasons (n=39) for not logging in to the MOPOrtal service at all were (1) not interested or laziness (51%), (2) forgot the service (49%), and (3) technical problems (15%). Among service users, the most common reasons reported for not logging were technical problems or discomfort with the wrist-worn physical activity monitor.

On the basis of the questionnaire (n=94), 90% of MOPOrtal users reported that data related to physical activity (diary and feedback) were important functionalities of the service. Additionally, instructions, test, and goals on physical activity (11%) and general information on health (11%) were also ranked as important functionalities in MOPOrtal. This was also supported by the log data showing that personal data on physical activity was the most used functionality in the service. The service users selected most often sports, movies, or music as their interests. On the other hand, camps, living, well-being, art, theater, or literature were more seldom selected as interests. Mostly, selected personal goals were increasing muscle mass and strength or stamina, whereas the weight control was the most seldom selected. Feedback graphs on daily and weekly physical activity motivated 65% of participants using MOPOrtal service. The reasons why these data did not motivate to move were that these persons were already physically active, they were not interested in physical activity, or they did not need motivation. Additionally, 61% found the feedback messages related to goals to be motivating for physical activity. Tips of the week, including physical activity and wellness messages, were evaluated to be mostly clear, interesting, and reliable.

Table 1. Baseline characteristics of the study participants (N=496). Values are mean (standard deviation) unless otherwise specified. N values for the baseline moderate to vigorous physical activity and sedentary time were 87 and 80 for the intervention and control groups, respectively.

Variable	Intervention (n=250)	Control (n=246)
Age, in years	17.9 (0.7)	17.8 (0.6)
Student, n (%)	218 (92.7)	214 (92.2)
Height, cm	177.9 (6.7)	178.1 (6.0)
Weight, kg	73.4 (15.0)	72.9 (14.0)
BMI[a], kg/m^2	23.2 (4.5)	23.0 (4.2)
Waist circumference, cm	81.9 (10.9)	81.9 (10.1)
Body fat, %	16.5 (8.5)	16.7 (8.3)
Muscle mass, %	47.1 (4.9)	46.9 (4.9)
Grip strength (mean), kg	45.6 (8.1)	45.6 (7.3)
Estimated maximal aerobic fitness, mL/min/kg	53.6 (7.3)	53.0 (6.8)
Daily moderate to vigorous physical activity[b] at baseline, min	59.6 (26.2)	61.9 (27.0)
Daily sedentary time at baseline, h	10.4 (2.1)	10.3 (1.9)
At the action or maintenance stage of physical activity[c] adaption, n (%)	167 (74.2)	152 (68.2)
Current smoker, n (%)	45 (19.6)	48 (21.3)
At least 6 servings of alcohol ≥ once a week, n (%)	43 (20.3)	43 (19.2)

[a]BMI: body mass index.

[b]MVPA: moderate to vigorous physical activity.

[c]PA: physical activity.

Some feedback for further development of MOPOrtal was obtained from the end questionnaire. More visual and clearer user interface (n=6 respondents), more interesting content (n=5), more simple (n=3), or technically more solid and mature solution (n=1) were suggested.

Those men who visited MOPOrtal at least once during the trial had a slightly higher body mass index (BMI) (mean difference of 1.2 kg/m^2, 95% CI 0.1 kg/m^2-2.2 kg/m^2) and body fat percentage (mean difference of 2.2%, 95% CI 0.1%-4.2%) at baseline compared with all other participants of the intervention group. Otherwise, there were no differences in anthropometry and fitness at baseline between these two groups.

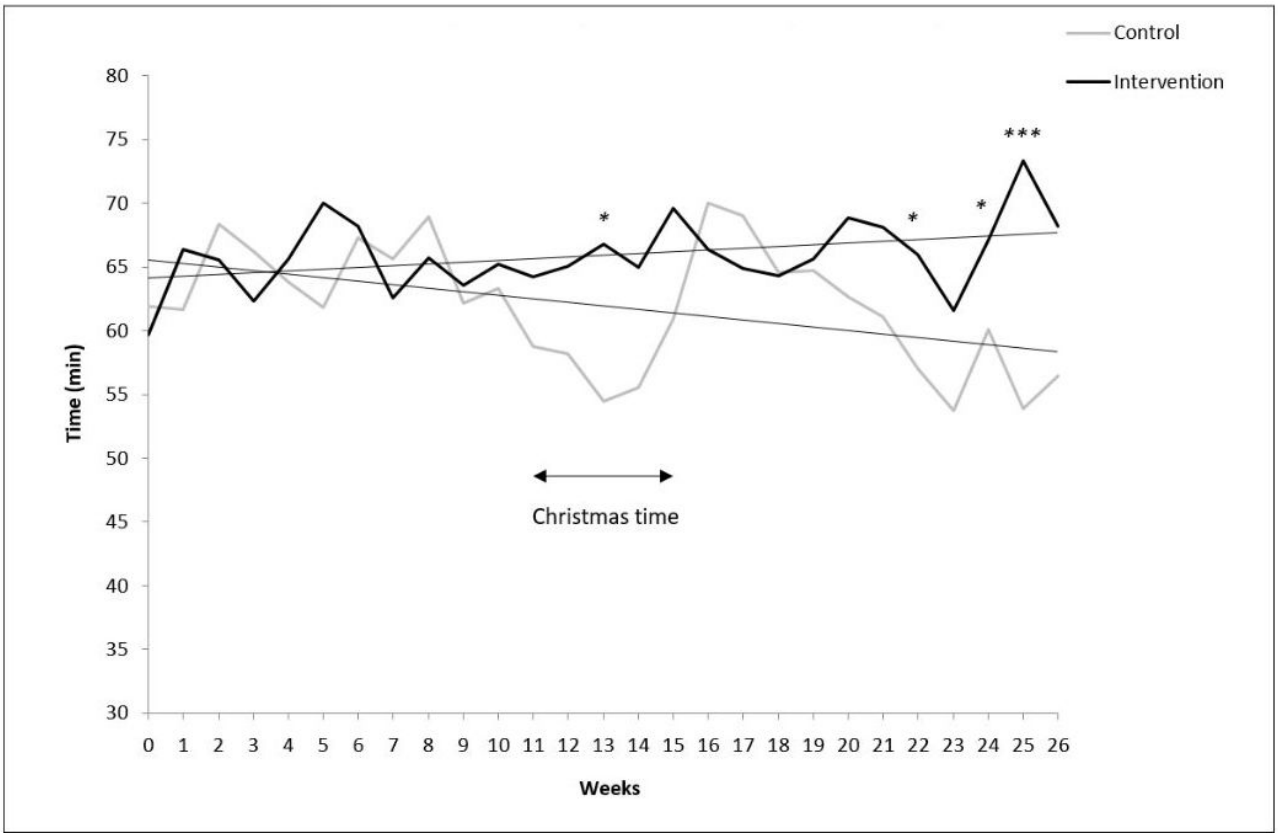
By the end of the trial, 178 men in the intervention group (95.2% of those who completed the study) had unlocked the screen of the activity monitor to show daily activity. The total number of valid days of objectively measured activity data was 15,364 (76.4% of all the uploaded data). Data were provided by 138 participants from the intervention group and 138 from the control group. At least 1 valid week was obtained from 120 (48.0%, 120/250) and 110 (44.7%, 110/246) participants in the intervention and control groups, respectively. The average number of valid weeks per person was 10 (median 8) in the intervention and 7 (median 4) in the control group, whereas the average daily usage time of the monitor was 15.1 hours and 15.7 hours in the intervention and control groups, respectively. Valid physical activity data were available from 87 participants in the intervention group and from 80 in the control group at the baseline week, and from 83/70, 56/43, 45/30, and 47/25 participants in the 6-week periods during the follow-up, respectively. Mean daily time occupied in sedentary behavior (1≤ MET< 2) at baseline was significantly higher in those participants who uploaded physical activity data until the end of the trial (n=73) compared with the participants who stopped to deliver data (n=94) during the trial (mean 10.7 hours vs 10.0 hours; P=.024, t test).

Effects of Mobile Service on Anthropometry, Fitness, and Physical Activity

Among completers (n=354), there was no statistically significant difference in the change in anthropometry and fitness measurements from baseline between the intervention and control groups. Within the intervention group, the change in the waist circumference differed between TTM-based inactive and active participants (mean change −0.3 cm vs 1.7 cm; P=.01, t test).

Among completers with valid physical activity data (n=167), there was a borderline difference in the change in mean daily time spent in moderate to vigorous physical activity between the intervention and control groups (11.9 min vs −9.1 min; P=.055, linear mixed model; Figure 4). During the last weeks, there was a significant difference between the groups in moderate to vigorous physical activity (P<.05 to P<.001, t test).

Figure 4. The mean daily time spent in moderate to vigorous physical activity for both study groups as measured by Polar Active during each week of the trial. The average standard deviation was 30.8 min and 24.4 min for the intervention and control groups, respectively. The weeks represent individual weeks from the baseline. Christmas holiday season took place during the study weeks 11 to 15 depending on when the individual started the trial. There was a borderline significant difference between the groups over time (P=.055, linear mixed model). MET: metabolic equivalent. * P<.05; *** P<.001.* P<0.05; *** P<0.001.



The intervention group succeeded to maintain their physical activity level over the Christmas season, whereas physical activity of the control group dropped during the holidays (P<.05, t test; Figure 4).

Among those men of the intervention group who logged on the service at least once during the trial, the usage frequency of the service was not associated with the change from baseline in the mean daily time spent in moderate to vigorous physical activity (ρ=.045, P=.77). Instead, within the intervention group, baseline vigorous physical activity was inversely associated with the change in daily moderate to vigorous physical activity time during the trial (R=−.382, P=.01).

Discussion

Principal Findings

In this 6-month trial, we assessed the feasibility of a fully automated, gamified, tailored Web-based mobile service (MOPOrtal) among young men. Additionally, we evaluated the effects of the service on anthropometry and fitness and objectively measured physical activity. The young men considered the various service functionalities related to physical activity important. The overall compliance was low, varying widely from low to moderate log and data upload frequency. Adherence to upload physical activity data was higher in those participants whose baseline sedentary time was higher. The mobile service had no effect on anthropometry or fitness during the 6-month trial, except reduced waist circumference in the most inactive men. Among the completers with valid physical activity data, there was a positive trend over time in favor of the intervention group in daily time spent in moderate to vigorous physical activity. Low amount of daily vigorous PA at baseline was found to be associated with the increase in moderate to vigorous physical activity during the trial.

The majority of MOPOrtal users perceived the functionalities related to physical activity important, motivating, and related to their personal goals. Both tailoring and gamification were applied to increase compliance by making the service more relevant, engaging, and interesting for the individual. However, the overall compliance was limited. Technical problems, in some degree immature user interface design, and fragmented functionalities were recognized as challenges for the perceived ease of use. To motivate those who are not interested in physical activity but might still benefit from physical activity information and guidance, more persuasive and behavior change–supporting intelligence should be implemented. Compatibility with a variety of sensors and devices would probably increase the usability and feasibility of the service. For successful design of mobile services in future, gamification should be an inseparable and coherent part of the integrity. Additional tailoring, for example, based on physical activity profiling [53], should also be taken into account.

To our knowledge, this is one of the first Web-based physical activity-related intervention studies implemented in a home

setting and including young male participants [23]. Unlike many previous studies, this study did not include face-to-face meetings except at baseline and at the end of the trial, allowing the evaluation of the impact based on the Web-based service only [12,22]. The 9-week intervention study conducted among adults with a fully automated Internet-based behavior change system, including continuous-time measurement of physical activity with a wrist-worn accelerometer, achieved consistently a 20 min difference per day in the time spent in moderate physical activity in the intervention group compared with the control group [16]. In addition, a Web-based intervention without any face-to-face meetings in older adults obtained similar results compared with this study, including 11-min increase in time spent in moderate to vigorous physical activity in the intervention group [20]. However, in addition to the difference in the intervention target group, the study duration was only a half, and objective measurement of physical activity was sampled in 7-day intervals at baseline and at 3 months [20].

In our previous pilot study, feedback from a wrist-worn activity monitor had a short-term positive effect on physical activity and sedentary behavior in young men [31]. Here, we observed a trend for a long-term effect in daily moderate to vigorous physical activity time, especially among those with low amount of vigorous physical activity at baseline. However, these data may be biased because of limited sample size, and we cannot distinguish whether the positive effect is a result of the gamified service or the feedback given by the wrist-worn physical activity monitor. The transition from adolescence to adulthood typically includes major life events and is an important phase to interfere with physical activity motivation to prevent negative changes in physical activity behavior and health in future [8]. We implemented different game mechanics in the MOPOrtal service to make it more engaging and attractive. MOPOrtal was a multicomponent service, and the impact of the game that was entered through a portal site cannot be distinguished from the results. However, additional improvements are needed to engage the user to maintain the interest to use the service for a longer time, especially in population-based studies in which the motivation level of the participants may vary. Although the annual military call-ups provide a large truly population-based study sample, this setting may have its challenges in recruiting participants for an eHealth trial, which requires highly motivated and active study participants to succeed [21]. Additionally, the high percentage of young men who declined to participate in the study during the call-ups showed a limited level of motivation among conscription-aged men.

The challenge with low usage of behavioral change services has been revealed also in earlier studies [54]. In several studies the use of an app or a service has dropped after the first month of the study [55]. In a study where a mobile phone app was used together with a face-to-face school-based program in adolescent boys, 20% of participants did not use the app at all [18]. The study using physical activity monitoring and a tailored physical activity coaching website for increasing physical activity reported that only 24% (n=10) of the participants had uploaded physical activity data regularly to the service during the 3-month trial [17]. In our study, physical activity data were provided to the service by 55.2% (138/250) of the participants in the intervention group, and from those, at least 20 weeks of data were obtained only from 21.7% (31/138; data not shown).

As the used behavior change model, that is, TTM, has been originally developed for changing unhealthy behavior [40], it can be discussed whether the model is suitable for the group in which 74.2% (n=167) of the participants are at the action or maintenance stage of physical activity adaptation already at the beginning of the study. The effect of the intervention might have been different if only inactive young men would had been recruited, which was supported by the different change in waist circumference found between inactive and active participants. In addition, at the beginning of the trial, daily objectives were not told, instead, the men from the intervention group achieved the information concerning physical activity recommendation through the MOPOrtal service and physical activity monitor. It is not known whether the information concerning physical activity objectives given by the staff who recruited participants to the trial at baseline would have increased or decreased the compliance of the study.

In this study, the anthropometry at baseline was related to the use of MOPOrtal. Participants who visited MOPOrtal at least once during the trial had a slightly higher BMI and body fat percentage at baseline compared with those who did not use the service at all. In addition, sedentary time at baseline was higher in those participants who uploaded physical activity data until at the end of the trial compared with the participants who stopped to deliver data during the trial. In the study by Tercyak et al, the presence of several behavioral risk factors, such as high BMI and insufficient physical activity, has been positively associated with willingness to use technology for health-promotion purposes among adolescents [56]. In addition, in this 6-month trial, we found out that the anthropometry and fitness at baseline were related to the attendance of the final measurements. Those young men who did not take part in the follow-up measurements had significantly higher body fat as well as lower estimated maximal aerobic fitness and grip strength at baseline compared with all other study participants (data not shown).

There was no change in anthropometry or fitness among the completers in the intervention group except for reduced waist circumference in the most inactive men. This might be because of the low adherence to service use, minor addition in daily moderate to vigorous physical activity time, or the relatively short duration of the trial. In some previous studies, a decrease in body weight and BMI following the use of a mobile phone app has been presented. However, the participants in the successful studies have been adults at increased risk of obesity [55]. A recent study assessing the efficacy of mobile phone technology for the treatment of obesity suggested that some level of counseling is needed in addition to the mobile phone app to improve anthropometry and fitness [57].

On the basis of the physical activity data recorded during this trial, the control group had a drop in moderate to vigorous physical activity during the holiday season in Christmas, followed by the highest peak immediately after Christmas. In contrast, the intervention group maintained their elevated physical activity level during the whole holiday season. In earlier

studies, in Western countries, the holiday season in December has been shown to have a negative effect on body fat in college students [58], but the effects of the holiday season on physical activity behavior are not known. However, physical activity behavior shows seasonal differences. After the summer, physical activity usually starts to decline, reaching its lowest level during the winter (January-March) [59,60]. In this study, the corresponding declining trend in time spent in moderate to vigorous physical activity can be seen among the control group, as the trial began in the autumn and ended at the end of the winter season.

Strengths and Limitations

The main strengths of this study were the large sample size and the population-based randomized controlled design. Another strength was the continuous measurement of physical activity within both study groups, allowing an objective assessment of the change in moderate to vigorous physical activity without any self-reported or user-entered data. The continuous measurement of physical activity also enabled both real-time and Web-based feedback to the intervention group. In addition, the home-based setting (without any face-to-face meetings except at baseline and at 6 months), instead of the more often used school setting, was another strength, allowing for better generalizability of the intervention [23].

One major limitation of this study was the missing physical activity data, the amount of which increased toward the end of the trial. Physical activity data needed to be uploaded to the database at least once every 3 weeks by the study participants, otherwise older data were overwritten by new data. Hence, it may be that some participants in the intervention group considered the feedback given by the PA monitor itself to be enough, and thus, they probably did not see a need for providing their physical activity data to the service. The stored data showed that those participants who did not use the MOPOrtal service, did not upload the physical activity data at all during the trial. More comprehensive self-monitoring with wearables could be obtained with wireless and automatic data transmission. Another limitation was the narrow age range, which limits the generalization of the results. Any long-term follow-up physical activity measurement after the end of the 6-month trial was not conducted in this study, which can also be seen as a limitation.

Conclusions

The various functionalities related to physical activity of the gamified tailored mobile service were considered important by the young men. However, the compliance to the service was limited. Adherence to upload activity data was higher in those participants whose baseline sedentary time was higher. Within the current setup, the mobile service had no effect on anthropometry or fitness during the 6-month trial, except reduced waist circumference in the most inactive men. Among completers with valid physical activity data, the trial had a borderline positive effect on moderate to vigorous physical activity, especially among those with low amount of vigorous physical activity at baseline. Further development is still needed to improve the feasibility and adherence of an integrated multifunctional service. Mobile services need to be further examined among populations of various ages.

Acknowledgments

This study was funded by the Ministry of Education and Culture (DNRO 125/627/2009, 98/627/2010, 97/627/2011); Centre for Military Medicine; Centre for Economic Development, Transport, and the Environment of North Ostrobothnia (European Social Fund, project number S11580); the Finnish Funding Agency for Technology and Innovation (European Regional Development Fund, 70037/2010 and 70035/2011), the Northern Ostrobothnia Hospital District; Infotech Oulu; the Juho Vainio Foundation; and the Finnish Cultural Foundation. The funders had no role in the planning or implementation of the intervention. The authors acknowledge Dr Paula Virtanen, Oulu Deaconess Institute Foundation, Polar Electro Ltd, Ludocraft Ltd, Valkee Ltd, the City of Oulu, Virpiniemi Sport Institute, and the Finnish Defence Forces for their support.

Conflicts of Interest

Riikka Ahola is currently an employee of Polar Electro. All other authors of this paper declare no conflicts of interest. Companies involved in the study had no role in the analyses and reporting.

Multimedia Appendix 1

CONSORT-EHEALTH v1.6 checklist.

[PDF File (Adobe PDF File), 7MB - mhealth_v5i10e146_app1.pdf]

References

1. Ekelund U, Luan J, Sherar LB, Esliger DW, Griew P, Cooper A, International Children's Accelerometry Database (ICAD) Collaborators. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. J Am Med Assoc 2012 Feb 15;307(7):704-712 [FREE Full text] [doi: 10.1001/jama.2012.156] [Medline: 22337681]
2. World Health Organization. Global recommendations on physical activity for health. Geneva, Switzerland: WHO Press; 2010:2015-2001.

3. Wen CP, Wai JP, Tsai MK, Yang YC, Cheng TY, Lee M, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 2011 Oct 1;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6] [Medline: 21846575]
4. Dencker M, Andersen LB. Accelerometer-measured daily physical activity related to aerobic fitness in children and adolescents. J Sports Sci 2011 Jun;29(9):887-895. [doi: 10.1080/02640414.2011.578148] [Medline: 21604226]
5. McGuire KA, Ross R. Incidental physical activity is positively associated with cardiorespiratory fitness. Med Sci Sports Exerc 2011 Nov;43(11):2189-2194. [doi: 10.1249/MSS.0b013e31821e4ff2] [Medline: 21502894]
6. Aarnio M, Winter T, Kujala U, Kaprio J. Associations of health related behaviour, social relationships, and health status with persistent physical activity and inactivity: a study of Finnish adolescent twins. Br J Sports Med 2002 Oct;36(5):360-364 [FREE Full text] [Medline: 12351335]
7. 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]
8. Telama R, Yang X. Decline of physical activity from youth to young adulthood in Finland. Med Sci Sports Exerc 2000 Sep;32(9):1617-1622. [Medline: 10994914]
9. Corder K, Sharp SJ, Atkin AJ, Griffin SJ, Jones AP, Ekelund U, et al. Change in objectively measured physical activity during the transition to adolescence. Br J Sports Med 2014 Apr 9;49(11):730-736 [FREE Full text] [doi: 10.1136/bjsports-2013-093190] [Medline: 24273308]
10. Wu S, Cohen D, Shi Y, Pearson M, Sturm R. Economic analysis of physical activity interventions. Am J Prev Med 2011 Feb;40(2):149-158 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.029] [Medline: 21238863]
11. 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):1155-62; quiz 1163. [Medline: 11678486]
12. Hamel LM, Robbins LB, Wilbur J. Computer- and web-based interventions to increase preadolescent and adolescent physical activity: a systematic review. J Adv Nurs 2011 Feb;67(2):251-268. [doi: 10.1111/j.1365-2648.2010.05493.x] [Medline: 21198800]
13. Pratt M, Sarmiento OL, Montes F, Ogilvie D, Marcus BH, Perez LG, Lancet Physical Activity Series Working Group. The implications of megatrends in information and communication technology and transportation for changes in global physical activity. Lancet 2012 Jul 21;380(9838):282-293. [doi: 10.1016/S0140-6736(12)60736-3] [Medline: 22818940]
14. Lenhart A. Pewinternet. Washington, DC: Pew Research Center; 2015. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [WebCite Cache ID 6ixtagpeP]
15. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
16. Hurling R, Catt M, Boni MD, 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;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
17. Slootmaker SM, Chinapaw MJ, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
18. Lubans DR, Smith JJ, Skinner G, Morgan PJ. Development and implementation of a smartphone application to promote physical activity and reduce screen-time in adolescent boys. Front Public Health 2014;2:42 [FREE Full text] [doi: 10.3389/fpubh.2014.00042] [Medline: 24904909]
19. Barwais FA, Cuddihy TF, Tomson LM. Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial. Health Qual Life Outcomes 2013;11:183 [FREE Full text] [doi: 10.1186/1477-7525-11-183] [Medline: 24168638]
20. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, De Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
21. Hansen AW, Grønbæk M, Helge JW, Severin M, Curtis T, Tolstrup JS. Effect of a Web-based intervention to promote physical activity and improve health among physically inactive adults: a population-based randomized controlled trial. J Med Internet Res 2012;14(5):e145 [FREE Full text] [doi: 10.2196/jmir.2109] [Medline: 23111127]
22. Lau PW, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
23. Ashton LM, Hutchesson MJ, Rollo ME, Morgan PJ, Collins CE. A scoping review of risk behaviour interventions in young men. BMC Public Health 2014;14:957 [FREE Full text] [doi: 10.1186/1471-2458-14-957] [Medline: 25224717]
24. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]

25. Blackman KC, Zoellner J, Kadir A, Dockery B, Johnson SB, Almeida FA, et al. Examining the feasibility of smartphone game applications for physical activity promotion in middle school students. Games Health J 2015 Oct;4(5):409-419. [doi: 10.1089/g4h.2014.0120] [Medline: 26287931]
26. Burke B. Gamify: how gamification motivates people to do extraordinary things. Brookline, MA: Bibliomotion; 2014.
27. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78. [doi: 10.1177/0141076813480996] [Medline: 23481424]
28. Tabak M, Dekker-van WM, van Dijk H, Vollenbroek-Hutten M. Promoting daily physical activity by means of mobile gaming: a review of the state of the art. Games Health J 2015 Dec;4(6):460-469. [doi: 10.1089/g4h.2015.0010] [Medline: 26397179]
29. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
30. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
31. Jauho A, Pyky R, Ahola R, Kangas M, Virtanen P, Korpelainen R, et al. Effect of wrist-worn activity monitor feedback on physical activity behavior: a randomized controlled trial in Finnish young men. Prev Med Rep 2015;2:628-634 [FREE Full text] [doi: 10.1016/j.pmedr.2015.07.005] [Medline: 26844128]
32. Ahola R, Pyky R, Jämsä T, Mäntysaari M, Koskimäki H, Ikäheimo TM, et al. Gamified physical activation of young men--a Multidisciplinary Population-Based Randomized Controlled Trial (MOPO study). BMC Public Health 2013;13:32 [FREE Full text] [doi: 10.1186/1471-2458-13-32] [Medline: 23311678]
33. Pyky R, Koivumaa-Honkanen H, Leinonen A, Ahola R, Hirvonen N, Enwald H, et al. Effect of tailored, gamified, mobile physical activity intervention on life satisfaction and self-rated health in young adolescent men: a population-based, randomized controlled trial (MOPO study). Comput Human Behav 2017 Jul;72:13-22 [FREE Full text] [doi: 10.1016/j.chb.2017.02.032]
34. Kinnunen H, Tanskanen M, Kyröläinen H, Westerterp KR. Wrist-worn accelerometers in assessment of energy expenditure during intensive training. Physiol Meas 2012 Nov;33(11):1841-1854. [doi: 10.1088/0967-3334/33/11/1841] [Medline: 23110981]
35. Brugniaux JV, Niva A, Pulkkinen I, Laukkanen RM, Richalet J, Pichon AP. Polar Activity Watch 200: a new device to accurately assess energy expenditure. Br J Sports Med 2010 Mar;44(4):245-249. [doi: 10.1136/bjsm.2007.045575] [Medline: 18413338]
36. Cain KL, Sallis JF, Conway TL, Van Dyck D, Calhoon L. Using accelerometers in youth physical activity studies: a review of methods. J Phys Act Health 2013 Mar;10(3):437-450. [Medline: 23620392]
37. Luoto T, Korpelainen R, Röning J, Ahola R, Enwald H, Hirvonen N, et al. Gamified persuasion: user experiences of online activation service. Int J Sociotechnology Knowl Dev 2014;6(4):1-17. [doi: 10.4018/ijskd.2014100101]
38. ludocraft. Clans of Oulu URL: http://www.ludocraft.com/education/clans-of-oulu/ [accessed 2017-09-13] [WebCite Cache ID 6tSDTnjz2]
39. Tuovinen L, Ahola R, Kangas M, Korpelainen R, Siirtola P, Luoto T, et al. Software Design Principles for Digital Behavior Change Interventions - Lessons Learned from the MOPO Study. 2016 Presented at: Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies; Feb ; Rome, Italy. 5; 2016; Rome, Italy p. 21-23. [doi: 10.5220/0005656101750182]
40. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [Medline: 6863699]
41. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
42. Marcus BH, Simkin LR. The transtheoretical model: applications to exercise behavior. Med Sci Sports Exerc 1994 Nov;26(11):1400-1404. [Medline: 7837962]
43. Woods C, Mutrie N, Scott M. Physical activity intervention: a transtheoretical model-based intervention designed to help sedentary young adults become active. Health Educ Res 2002 Aug;17(4):451-460 [FREE Full text] [Medline: 12197590]
44. Nigg CR, Geller KS, Motl RW, Horwath CC, Wertin KK, Dishman RK. A research agenda to examine the efficacy and relevance of the transtheoretical model for physical activity behavior. Psychol Sport Exerc 2011 Jan;12(1):7-12 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.004] [Medline: 21113323]
45. Hirvonen N, Enwald H, Bath PA, Pyky R, Korpelainen R, Huotari ML. Individual factors affecting preferences for feedback message tactics in the contexts of physical activity. J Health Commun 2015;20(2):220-229. [doi: 10.1080/10810730.2014.925015] [Medline: 25491473]
46. Enwald H, Hirvonen N, Korpelainen R, Huotari ML. Young men´s perceptions of fear appeal versus neutral health messages ? associations with everyday health information literacy, education, and health. 2015 Presented at: Proceedings of ISIC, the Information Behaviour Conference. Part 2. Information Research, 20(1); 2014; Leeds, United Kingdom p. 15-19.
47. Cardinal BJ. The stages of exercise scale and stages of exercise behavior in female adults. J Sports Med Phys Fitness 1995 Jun;35(2):87-92. [Medline: 7500632]

48. ukkinstituutti.fi. Helsinki: Ministry of Education and Culture & Nuori Suomi ry; 2008. Recommendation for the physical activity of school-aged children [Finnish] URL: http://www.ukkinstituutti.fi/filebank/1477-Fyysisen_aktiivisuuden_suositus_kouluikaisille.pdf [accessed 2017-09-25] [WebCite Cache ID 6tjq7Dzjj]
49. Bohannon R. Hand-held dynamometry: a practicable alternative for obtaining objective measures of muscle strength. Isokinetics Exerc Sci 2012;20(4):301-315. [doi: 10.3233/IES-2012-0476]
50. Väinämö K, Nissilä S, Mäkikallio T, Tulppo M, Röning J. Artificial neural networks for aerobic fitness approximation. In: Proceedings of the International Conference on Neural Networks.: IEEE; 1996 Presented at: International Conference on Neural Networks (ICNN); June 1996; Washington DC, USA p. 1939-1944. [doi: 10.1109/ICNN.1996.549198]
51. Crumpton S, Williford HN, O'Mailia S, Olson MS, Woolen LE. Validity of the Polar M52 Heart Rate Monitor in Predicting VO2max. Med Sci Sports Exerc 2003;35:S193. [doi: 10.1097/00005768-200305001-01078]
52. Peltola K, Hannula M, Held T, Kinnunen H, Nissilä S, Laukkanen R, et al. Validity of polar fitness test based on heart rate variability in assessing VO2max in trained individuals. 2000 Presented at: Proceedings of the 5th Annual Congress of the European College of Sport Science; 2000; Jyväskylä, Finland.
53. Pyky R, Jauho A, Ahola R, Ikäheimo TM, Koivumaa-Honkanen H, Mäntysaari M, et al. Profiles of sedentary and non-sedentary young men - a population-based MOPO study. BMC Public Health 2015 Nov 23;15:1164 [FREE Full text] [doi: 10.1186/s12889-015-2495-6] [Medline: 26596355]
54. Kohl LF, Crutzen R, De Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]
55. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
56. Tercyak KP, Abraham AA, Graham AL, Wilson LD, Walker LR. Association of multiple behavioral risk factors with adolescents' willingness to engage in eHealth promotion. J Pediatr Psychol 2009 Jun;34(5):457-469 [FREE Full text] [doi: 10.1093/jpepsy/jsn085] [Medline: 18723566]
57. Allen JK, Stephens J, Dennison HC, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
58. Hull HR, Hester CN, Fields DA. The effect of the holiday season on body weight and composition in college students. Nutr Metab (Lond) 2006;3:44 [FREE Full text] [doi: 10.1186/1743-7075-3-44] [Medline: 17192197]
59. Merchant AT, Dehghan M, Akhtar-Danesh N. Seasonal variation in leisure-time physical activity among Canadians. Can J Public Health 2007;98(3):203-208. [Medline: 17626385]
60. Pivarnik JM, Reeves MJ, Rafferty AP. Seasonal variation in adult leisure-time physical activity. Med Sci Sports Exerc 2003 Jun;35(6):1004-1008. [doi: 10.1249/01.MSS.0000069747.55950.B1] [Medline: 12783049]

Abbreviations

BMI: body mass index
GPS: global positioning system
MET: metabolic equivalent
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SEE: standard error of estimate
TTM: transtheoretical model of behavior change

Edited by G Eysenbach; submitted 27.09.16; peer-reviewed by E Lyons, M Duncan, A McDougall, K Thorsteinsen; comments to author 07.01.17; revised version received 29.05.17; accepted 15.08.17; published 10.10.17.

Please cite as:
Leinonen AM, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, Enwald H, Ikäheimo TM, Röning J, Keinänen-Kiukaanniemi S, Mäntysaari M, Korpelainen R, Jämsä T
Feasibility of Gamified Mobile Service Aimed at Physical Activation in Young Men: Population-Based Randomized Controlled Study (MOPO)
JMIR Mhealth Uhealth 2017;5(10):e146
URL: https://mhealth.jmir.org/2017/10/e146/
doi:10.2196/mhealth.6675
PMID:29017991

Original Paper

Willingness to Use Mobile Phone Apps for HIV Prevention Among Men Who Have Sex with Men in London: Web-Based Survey

William C Goedel[1], BA; Jason W Mitchell[2], PhD; Paul Krebs[1], PhD; Dustin T Duncan[1], ScD

[1]Department of Population Health, School of Medicine, New York University, New York, NY, United States

[2]Office of Public Health Studies, University of Hawaii at Manoa, Honolulu, HI, United States

Corresponding Author:
William C Goedel, BA
Department of Population Health
School of Medicine
New York University
227 East 30th Street
New York, NY,
United States
Phone: 1 6465012715
Fax: 1 6465012706
Email: wcg219@nyu.edu

Abstract

Background: Many men who have sex with men (MSM) use apps to connect with and meet other MSM. Given that these apps are often used to arrange sexual encounters, it is possible that apps may be suitable venues for messages and initiatives related to HIV prevention such as those to increase HIV testing rates among this population.

Objective: The purpose of this study was to assess willingness to use a new app for reminders of when to be tested for HIV infection among a sample of MSM in London who use apps to arrange sexual encounters.

Methods: Broadcast advertisements targeted users of a popular social-networking app for MSM in London. Advertisements directed users to a Web-based survey of sexual behaviors and sexual health needs. Willingness to use apps for reminders of when to be tested for HIV was assessed. In addition, participants responded to items assessing recent sexual behaviors, substance use, and demographic characteristics. Exploratory analyses were undertaken to examine differences in willingness to use an app by demographic and behavioral characteristics.

Results: Broadcast advertisements yielded a sample of 169 HIV-negative MSM. Overall, two-thirds (108/169, 63.9%) reported willingness to use an app to remind them when to be tested for HIV. There were no significant differences in willingness to use these apps based on demographic characteristics, but MSM who reported recent binge drinking and recent club drug use more frequently reported willingness to use this app compared to their nonusing counterparts.

Conclusions: MSM in this sample are willing to use a new app for HIV testing reminders. Given the high levels of willingness to use them, these types of apps should be developed, evaluated, and made available for this population.

(JMIR Mhealth Uhealth 2017;5(10):e153) doi:10.2196/mhealth.8143

KEYWORDS

mobile phone apps; mHealth; men who have sex with men (MSM); HIV

Introduction

The human immunodeficiency virus (HIV) epidemic in the United Kingdom impacts gay, bisexual, and other men who have sex with men (MSM) at a disproportionate rate compared to the general population. MSM are estimated to represent 3.3% of the male population in the United Kingdom but represent between 66.7% and 69.0% of all males living with HIV infection as of 2015 and 73.0% of newly diagnosed HIV infections among males in 2015 [1]. It is estimated that between 7.2% and 18.9% of HIV-infected MSM are unaware of their infection status and that 30.0% of newly diagnosed HIV-infected MSM are diagnosed “late” (ie, with a CD4 count of less than 350 cells per cubic millimeter within 3 months of diagnosis) [1]. Regular HIV testing has a number of important benefits for MSM who test positive and who test negative. For MSM who test negative,

HIV testing provides the opportunity for risk reduction counseling to encourage behavior change that reduces risk for HIV infection and allows the individual to remain HIV negative [2]. For MSM who test positive, HIV testing improves early awareness of infection status and linkage to care [3]. New interventions to increase HIV testing frequency among MSM in the United Kingdom are needed to further identify new cases of HIV infection and effectively link them to treatment and to reduce engagement in risk behaviors by HIV-uninfected and HIV-infected MSM.

For the past decade, the Internet has been used to deliver and disseminate HIV prevention interventions to diverse populations, including gay, bisexual, and other MSM [4]. Frequently, these interventions have included virtual scenarios and simulations, decision making with virtual characters, and detailed answers or feedback following knowledge tests. For example, Health mPowerment, an intervention based on the Institute of Medicine's integrated model of behavior theory aimed at increasing HIV and sexually transmitted infection (STI) knowledge among young black MSM, incorporates live chats with outreach workers, quizzes, personalized health and journals of sexual activities, and decision support tools for assessing risk behaviors on a single website [5]. Building on these Internet-based interventions and keeping with the pace of technological developments, mobile phone–based HIV prevention interventions are currently in various stages of development and testing. Many of these mobile phone–based interventions incorporate text messaging delivery and have demonstrated some success. For example, Bourne and colleagues [6] found that MSM who enrolled in a text message–based intervention involving reminders for HIV/STI testing tailored based on participant risk behaviors and ability to return for testing were over 4 times more likely to retest for HIV and other STIs compared to those who did not receive the intervention.

Given the pervasiveness of mobile phones and associated apps [7], it may be best to incorporate and adapt aspects of these previously developed Internet- and text message–based interventions for delivery via apps, but it should be noted that their effectiveness is dependent on their continual use, as many individuals download health-related apps and discontinue using them for a variety of reasons [8]. The use of geosocial-networking apps, those that use Global Positioning System (GPS) technology to facilitate various types of connections between MSM, is common among MSM [9-11]. There is no clear consensus in the extant literature regarding the association between meeting sexual partners online or through apps and increased HIV risk behaviors [12]. It has been hypothesized that these digital venues may facilitate multiple partnerships and other HIV risk behaviors for MSM who already engage in these behaviors rather than act as a catalyst for these behaviors themselves [12]. However, while social and sexual network apps are accepted by MSM for HIV prevention and sexual health messaging [13], some companies who produce these geosocial-networking apps are ambiguous in their willingness to support HIV prevention programs [14].

Therefore, the development of new apps for HIV prevention purposes targeted to those potentially high-risk MSM who already use apps to meet sexual partners may be beneficial, but little is known about MSM willingness to engage with new app-related HIV prevention services. As such, the purpose of this study was to examine the willingness to use apps for HIV prevention purposes among a sample of MSM in London recruited from a popular geosocial-networking app used by MSM to meet sexual partners. In exploratory analyses, we examined differences in willingness to use these apps by demographic characteristics and recent sexual and substance use behaviors. This study specifically focused on London because it is an urban epicenter with a high prevalence of HIV among MSM [15] and a high number of MSM who use apps to meet sexual partners [16].

Methods

Sample Recruitment

Broadcast advertisements were placed on a popular geosocial-networking app commonly used by MSM to meet sexual partners in January 2016. These broadcasts were targeted specifically to the London metropolitan area. Consistent with recent research among MSM who use apps to meet sexual partners [9], users were shown an advertisement the first time they logged in to the app during four 24-hour periods [17]. The advertisement contained text encouraging users to click through to complete an anonymous Web-based survey on the health of gay and bisexual men in the United Kingdom. Participants were told that completing the survey would enter them in a lottery for a chance to earn £50 (US $67). At the end of the recruitment period, 1410 users had clicked through the advertisement and 202 users completed the survey, representing a cooperation rate of 14.33%. All protocols were approved via institutional review board prior to data collection.

Survey Measures

Willingness to Use Mobile Phone Apps for HIV Prevention

Respondents were asked, "How willing would you be to use an app to remind yourself when to get tested for HIV?" The wording of this potential intervention was selected given its alignment with the goals of primary prevention of HIV infection [18]. Using a 5-point Likert scale, response options for this item ranged from very willing to very unwilling. For analytical purposes, these categories were collapsed into willing, undecided, and unwilling.

Geosocial-Networking Mobile Phone App Use

Given that previous research has shown that many individuals who use health-related apps download them and then discontinue using them [8], all respondents were asked, "In the past 12 months, how many times have you deleted and re-downloaded Grindr?" with the aim of understanding consistency of use of geosocial-networking apps to connect with other MSM. Response options for this item included 8 or more times, 6 or 7 times, 4 or 5 times, 2 or 3 times, 1 time, and 0 times.

Recent Sexual Behaviors

All respondents were asked to report the total number of partners with whom they had insertive and receptive anal intercourse with and without a condom in the preceding 3 months. These

counts were then dichotomized (0 partners vs 1 or more partners).

Recent Substance Use Behaviors

Respondents were asked to select from a list of substances to indicate any use in the preceding 3 months, included alcohol (≥5 drinks in 1 sitting), cocaine, ecstasy, gamma-hydroxybutyric acid or gamma-butyrolactone (GHB/GBL), inhalant nitrites, ketamine, lysergic acid diethylamide (LSD), marijuana, and methamphetamine. Each of these substances was considered individually and a composite variable was created for club drug use (eg, use of ecstasy, GHB/GBL, ketamine, and/or LSD).

Demographic Variables

Respondents reported their age, sexual orientation, ethnic group membership, employment status, relationship status, and HIV status. Age was measured continuously in years and categorized as aged 18 to 24 years, 25 to 30 years, 31 to 40 years, 41 to 50 years, and 51 years and older. Sexual orientation was categorized as gay, bisexual, straight, and other. Ethnic group membership was categorized as white, black, Asian, and mixed/multiple ethnic groups/other. Employment status was categorized as employed, unemployed, retired, or student. Relationship status was categorized as being in a primary relationship with another man or not. HIV status was assessed based on self-report as negative or positive.

Statistical Analyses

Broadcast advertisements yielded an overall sample of 202 respondents within 96 hours. A total of 10 individuals who did not report their HIV status and 23 who self-reported as HIV-positive were excluded, restricting the analytical sample to 169 respondents. Descriptive statistics were calculated for all variables. Differences in willingness to use apps for HIV prevention purposes by demographic variables, recent substance use, and categorical measures of recent sexual behaviors were assessed using chi-square tests of independence. Significance was determined as $P<.05$.

Results

Sample Demographics

The average age of the sample was 36.7 (SD 11.5) years. A majority (165/169, 97.9%) self-identified as gay or bisexual. Almost three-fourths (122/169, 72.4%) identified their ethnic group as white or white British. A large percentage (144/169, 85.5%) were either currently employed or enrolled in school. Approximately one-fifth (32/169, 18.8%) reported currently being in a relationship with another man.

Recent Sexual Behaviors

Three-quarters of respondents (127/169, 75.2%) engaged in insertive anal intercourse in the preceding 3 months. Among respondents who had engaged in insertive anal intercourse, 59.9% (76/127) engaged in condomless insertive anal intercourse with 1 or more partners. Almost two-thirds of respondents (110/169, 64.9%) engaged in receptive anal intercourse in the preceding 3 months. Among respondents who had engaged in receptive anal intercourse, 57.3% (63/110) engaged in condomless receptive anal intercourse with 1 or more partners. About two-fifths of respondents (72/169, 42.6%) had engaged in a group sex event in the preceding 3 months.

Recent Substance Use Behaviors

Two-fifths of respondents (65/169, 38.6%) reported having had 5 or more drinks containing alcohol in 1 sitting in the preceding 3 months. With regard to other substances, 36.6% (62/169) reported using inhalant nitrites, 23% (39/169) reported using club drugs, 19.8% (33/169) reported using marijuana, 14.4% (24/169) reported using cocaine or crack cocaine, and 13.9% (23/169) reported using methamphetamine in the preceding 3 months.

Geosocial-Networking Mobile Phone App Use

About half of respondents (82/169, 48.5%) had not deleted Grindr in the preceding year. Among those who had deleted and re-downloaded Grindr in the preceding year, 69.2% (57/82) had deleted and re-downloaded the app 3 or fewer times.

Willingness to Use Mobile Phone Apps for HIV Prevention

Descriptive statistics are displayed in Table 1. Among the sample, 63.9% (108/169) were willing to use an app to remind them to get tested for HIV, 24.9% (42/169) were undecided, and 11.2% (19/169) were unwilling.

Exploratory analyses examining differences in self-reported willingness to use an app for HIV prevention by demographic characteristics (eg, age, ethnic group membership) and recent sexual (eg, engagement in condomless insertive anal intercourse) and substance use (eg, alcohol use, club drug use) behaviors are shown in Table 1. Respondents who reported binge drinking ($P=.03$) and club drug use ($P=.04$) more frequently reported more willingness to use an app for HIV testing reminders than their nonusing counterparts.

Table 1. Willingness to use an app for HIV testing reminders among HIV-negative men who have sex with men recruited from a popular social-networking app in London (n=169).

Demographic/behavioral characteristic	Frequency, n (%)	Rated very willing/willing, n (%)	*P* value
Age, years			.69
18 to 24	30 (16.8)	18 (69.3)	
25 to 30	46 (25.7)	24 (54.6)	
31 to 40	55 (30.7)	36 (70.6)	
41 to 50	30 (16.8)	17 (56.6)	
51 and older	18 (10.1)	13 (72.2)	
Sexual orientation			.80
Gay	156 (87.2)	92 (62.4)	
Bisexual	19 (10.6)	13 (76.5)	
Other	4 (2.2)	2 (66.7)	
Ethnic group membership			.99
White/white British	128 (71.5)	80 (65.6)	
Black/black British	10 (5.6)	6 (66.6)	
Asian/Asian British	14 (7.8)	6 (54.6)	
Mixed/multiple ethnic groups/other	25 (14.0)	14 (56.0)	
Employment status			.33
Employed	133 (74.3)	80 (62.6)	
Unemployed	17 (9.5)	9 (60.0)	
Student	23 (12.8)	13 (75.0)	
Retired	5 (2.8)	5 (100.0)	
Primary relationship with man			.97
Yes	32 (17.9)	18 (62.0)	
No	147 (82.1)	90 (64.2)	
Recent sexual behaviors			
Condomless receptive anal intercourse	62 (34.6)	39 (69.6)	.09
Condomless insertive anal intercourse	76 (42.5)	47 (65.8)	.50
Recent substance use			
Alcohol (≥5 drinks in a row)	71 (397)	48 (71.6)	.03
Cocaine	26 (14.5)	15 (62.5)	.18
Club drugs	40 (22.3)	30 (78.9)	.04
Inhalant nitrites	63 (35.2)	43 (70.5)	.69
Marijuana	38 (21.2)	24 (66.7)	.94
Methamphetamine	21 (11.7)	18 (90.0)	.12

Discussion

Principal Findings

This is the first study to examine the acceptability of the use of mobile phone apps for HIV testing reminders among a sample of MSM in London who use geosocial-networking apps to meet other MSM. Overall, respondents reported a willingness to use an app to remind them to be periodically tested for HIV (108/169, or 63.9%, rated themselves as willing or very willing). Our findings are similar to those observed among a sample of young MSM in Southern California, where about 70.0% (137/195) were willing to participate in an HIV prevention program delivered via an app [19]. In addition, we found no significant differences in willingness to use these apps based on demographic characteristics or sexual behaviors. The lack of significant variations in willingness to use these apps across most subgroups in this sample of MSM suggests that these intervention tools can be implemented in diverse populations of MSM living in London. However, the presence of significant differences among MSM who reported binge drinking and club

drug use may suggest a need for targeted prevention-related apps for these populations so that they can manage their HIV-related risks and substance use-related risks concurrently.

Given the high prevalence of inconsistent use of geosocial-networking apps to meet sexual partners in the preceding 12 months (where 87/169, or 51.5%, reported deleting and re-downloading the app at least once), the implementation of these interventions should include strategies to encourage consistency of use. Recent formative qualitative research has aimed at assessing factors associated with consistency of use. Among samples of HIV-negative MSM in Miami and Minneapolis, reliability, ease of use, and frequency of updates were cited as factors associated with whether an individual would keep and continue to use an app over time while poor performance and functionality and lack of use were primary reasons why MSM would delete an app from their phone [20]. The item used in this study examines consistency of use of social-networking apps, and this may not reflect potential consistency of use of health-related apps or HIV prevention-related apps. However, it is possible that HIV prevention-related features of app-based interventions may be more consistently used if integrated into an existing social-networking platform.

Future research is needed to identify and test potential apps to be used for HIV testing reminders among MSM who use geosocial-networking apps to meet sexual partners, particularly in contexts other than the United States, given that this is where most of the previous research has been conducted [21,22]. However, it is possible that the features of apps desired by MSM in the United States may be similar to those desired by MSM elsewhere. Previous research with HIV-negative MSM has suggested that apps for HIV prevention should encourage regular HIV testing by providing feedback on test reminders, tailored testing interval recommendations, and HIV test site-locating features [20]. In addition, these apps should be designed to be discrete in nature, protect privacy, and not necessarily appear overtly to be related to HIV prevention. Should another individual, for example, see an app explicitly for HIV prevention, it could expose an individual to HIV-related stigma (if others assume they are HIV-positive) and discourage them from using these types of apps.

Limitations

These findings are not without limitations. These findings are derived from data collected as part of a convenience sample of MSM from a single popular geosocial-networking app in a single metropolitan center in Western Europe. As such, these findings are likely not generalizable to broader populations of MSM on other apps or outside of these apps. These findings are also likely biased by some degree of self-selection, where those who are willing to engage with communication regarding HIV prevention via apps may have been more likely to participate. In addition, our measures of risk behaviors (eg, sexual and substance use behaviors) are crude, and there may be some misclassification with regard to behavioral risk. Future research should use more nuanced measures of these behaviors (eg, number of events of condomless receptive anal intercourse with a serodiscordant or unknown HIV status partner in a recall period, number of times using methamphetamine in a recall period) to further our understanding of the behavioral risk profiles of MSM who are willing to use apps for HIV prevention.

Conclusion

The majority of respondents in this sample of MSM who use apps to meet other MSM reported willingness to use an app to remind them to be tested for HIV infections. Apps to be used for HIV prevention interventions, such as those aimed at increasing HIV testing rates among MSM (particularly those who are already using apps to meet other MSM), should be developed, evaluated, and implemented.

Conflicts of Interest

None declared.

References

1. Public Health England. HIV in the United Kingdom. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/602942/HIV_in_the_UK_report.pdf [accessed 2017-09-28] [WebCite Cache ID 6tp2E2y2O]
2. Wolitski RJ, MacGowan RJ, Higgins DL, Jorgensen CM. The effects of HIV counseling and testing on risk-related practices and help-seeking behavior. AIDS Educ Prev 1997 Jun;9(3 Suppl):52-67. [Medline: 9241398]
3. Ulett KB, Willig JH, Lin H, Routman JS, Abroms S, Allison J, et al. The therapeutic implications of timely linkage and early retention in HIV care. AIDS Patient Care STDS 2009 Jan;23(1):41-49 [FREE Full text] [doi: 10.1089/apc.2008.0132] [Medline: 19055408]
4. Ybarra ML, Bull SS. Current trends in Internet- and cell phone-based HIV prevention and intervention programs. Curr HIV/AIDS Rep 2007 Dec;4(4):201-207. [Medline: 18366952]
5. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011 Feb;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
6. Bourne C, Knight V, Guy R, Wand H, Lu H, McNulty A. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. Sex Transm Infect 2011 Apr;87(3):229-231. [doi: 10.1136/sti.2010.048397] [Medline: 21296796]

7. Smith A. US smartphone use in 2015. Washington: Pew Internet and American Life Project URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2015/03/PI_Smartphones_0401151.pdf [accessed 2017-09-28] [WebCite Cache ID 6tp2OWADW]
8. 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]
9. Goedel WC, Duncan DT. Geosocial-networking app usage patterns of gay, bisexual, and other men who have sex with men: survey among users of Grindr, a mobile dating app. JMIR Public Health Surveill 2015;1(1):e4 [FREE Full text] [doi: 10.2196/publichealth.4353] [Medline: 27227127]
10. Lehmiller JJ, Ioerger M. Social networking smartphone applications and sexual health outcomes among men who have sex with men. PLoS One 2014;9(1):e86603 [FREE Full text] [doi: 10.1371/journal.pone.0086603] [Medline: 24466166]
11. Landovitz RJ, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]
12. Zou H, Fan S. Characteristics of men who have sex with men who use smartphone geosocial networking applications and implications for HIV interventions: a systematic review and meta-analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
13. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and feasibility of using established geosocial and sexual networking mobile applications to promote HIV and STD testing among men who have sex with men. AIDS Behav 2015 Mar;19(3):543-552. [doi: 10.1007/s10461-014-0942-5] [Medline: 25381563]
14. Chan PA, Towey C, Poceta J, Rose J, Bertrand T, Kantor R, et al. Online hookup sites for meeting sexual partners among men who have sex with men in Rhode Island, 2013: a call for public health action. Public Health Rep 2016 Mar;131(2):264-271 [FREE Full text] [doi: 10.1177/003335491613100210] [Medline: 26957661]
15. Desai S, Croxford S, Brown AE, Mitchell H, Hughes G, Delpech V. An overview of the HIV epidemic among men who have sex with men in the United Kingdom, 1999-2013. Euro Surveill 2015 Apr 09;20(14) [FREE Full text] [Medline: 25884147]
16. Grindr. URL: https://www.grindr.com/ [accessed 2017-09-28] [WebCite Cache ID 6tp30s3y9]
17. Duncan DT, Goedel WC, Mayer KH, Safren SA, Palamar JJ, Hagen D, et al. Poor sleep health and its association with mental health, substance use, and condomless anal intercourse among gay, bisexual, and other men who have sex with men. Sleep Health 2016 Dec;2(4):316-321. [doi: 10.1016/j.sleh.2016.07.003]
18. Kelly JA, Kalichman SC. Behavioral research in HIV/AIDS primary and secondary prevention: recent advances and future directions. J Consult Clin Psychol 2002 Jun;70(3):626-639 [FREE Full text] [Medline: 12090373]
19. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
20. Mitchell JW, Torres MB, Joe J, Danh T, Gass B, Horvath KJ. Formative work to develop a tailored HIV testing smartphone app for diverse, at-risk, HIV-negative men who have sex with men: a focus group study. JMIR Mhealth Uhealth 2016 Nov 16;4(4):e128 [FREE Full text] [doi: 10.2196/mhealth.6178] [Medline: 27852558]
21. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
22. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]

Abbreviations

GBL: gamma-butyrolactone
GHB: gamma-hydroxybutyric acid
GPS: Global Positioning System
HIV: human immunodeficiency virus
LSD: lysergic acid diethylamide
MSM: men who have sex with men
STI: sexually transmitted infection

Edited by C Dias; submitted 02.06.17; peer-reviewed by J Uhrig, C Shen; comments to author 21.06.17; revised version received 18.07.17; accepted 13.08.17; published 11.10.17.

Please cite as:
Goedel WC, Mitchell JW, Krebs P, Duncan DT
Willingness to Use Mobile Phone Apps for HIV Prevention Among Men Who Have Sex with Men in London: Web-Based Survey
JMIR Mhealth Uhealth 2017;5(10):e153
URL: http://mhealth.jmir.org/2017/10/e153/
doi:10.2196/mhealth.8143
PMID:29021132

Original Paper

Desire to Be Underweight: Exploratory Study on a Weight Loss App Community and User Perceptions of the Impact on Disordered Eating Behaviors

Elizabeth Victoria Eikey[1], PhD; Madhu C Reddy[2], PhD; Kayla M Booth[3], BA; Lynette Kvasny[4], PhD; Johnna L Blair[4], BS; Victor Li[5], MS; Erika S Poole[6], PhD

[1]Donald Bren School of Information and Computer Sciences, Department of Informatics, University of California, Irvine, Irvine, CA, United States
[2]Department of Communication Studies, Northwestern University, Evanston, IL, United States
[3]School of Computing and Information, University of Pittsburgh, Pittsburgh, PA, United States
[4]College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, United States
[5]University of Washington, Seattle, WA, United States
[6]Healthwise, Boise, ID, United States

Corresponding Author:
Elizabeth Victoria Eikey, PhD
Donald Bren School of Information and Computer Sciences
Department of Informatics
University of California, Irvine
6095 Donald Bren Hall
Irvine, CA, 92697-3440
United States
Phone: 1 9494381337
Email: eikeye@uci.edu

Abstract

Background: Mobile health (mHealth) apps for weight loss (weight loss apps) can be useful diet and exercise tools for individuals in need of losing weight. Most studies view weight loss app users as these types of individuals, but not all users have the same needs. In fact, users with disordered eating behaviors who desire to be underweight are also utilizing weight loss apps; however, few studies give a sense of the prevalence of these users in weight loss app communities and their perceptions of weight loss apps in relation to disordered eating behaviors.

Objective: The aim of this study was to provide an analysis of users' body mass indices (BMIs) in a weight loss app community and examples of how users with underweight BMI goals perceive the impact of the app on disordered eating behaviors.

Methods: We focused on two aspects of a weight loss app (DropPounds): profile data and forum posts, and we moved from a broader picture of the community to a narrower focus on users' perceptions. We analyzed profile data to better understand the goal BMIs of all users, highlighting the prevalence of users with underweight BMI goals. Then we explored how users with a desire to be underweight discussed the weight loss app's impact on disordered eating behaviors.

Results: We found three main results: (1) no user (regardless of start BMI) starts with a weight gain goal, and most users want to lose weight; (2) 6.78% (1261/18,601) of the community want to be underweight, and most identify as female; (3) users with underweight BMI goals tend to view the app as positive, especially for reducing bingeing; however, some acknowledge its role in exacerbating disordered eating behaviors.

Conclusions: These findings are important for our understanding of the different types of users who utilize weight loss apps, the perceptions of weight loss apps related to disordered eating, and how weight loss apps may impact users with a desire to be underweight. Whereas these users had underweight goals, they often view the app as helpful in reducing disordered eating behaviors, which led to additional questions. Therefore, future research is needed.

(JMIR Mhealth Uhealth 2017;5(10):e150) doi:10.2196/mhealth.6683

KEYWORDS

mobile applications; health apps; feeding and eating disorder; disordered eating behaviors; desire to be underweight; body mass index; weight loss; smartphone; mHealth; online community; forum; profile; posts; human computer interaction

Introduction

Motivation

While it is estimated that 30 million people in the United States have an eating disorder, many more have disordered eating behaviors, especially young women [1-3]. Research has shown that people with eating disorders and disordered eating behaviors utilize various forms of technology, including forums, websites, blogs, and social media, to support their recovery process or maintain the symptoms of their disorder [4-6]. However, only a few studies have been conducted on the use of mobile health (mHealth) apps for weight loss (weight loss apps) by users with eating disorders and disordered eating behaviors, despite their popularity [7].

Although many researchers have studied weight loss apps [8-13], few have studied weight loss apps in relation to disordered eating behaviors [14,15]. Most research on weight loss apps focuses on the use and design of these apps to encourage overall wellness, healthy diet, weight loss, and physical activity. In terms of eating disorders and disordered eating, researchers have begun to consider the use of weight loss apps [14,15], but studies to date do not highlight the prevalence of users with disordered eating behaviors within weight loss app communities.

Many studies on the role of technology for disordered eating behaviors focus on content that is in support of eating disorders or pro-eating disorder content. Pro-eating disorder content often considers eating disorders as a lifestyle choice rather than a disorder requiring treatment. Eating disorder–related material can be found on websites and social media platforms such as Facebook, Twitter, and Instagram [15-18], which are now often accessed via apps. Our work differs from prior research in that we focus on an app intended for weight loss instead of technology designed specifically for eating disorders or general-purpose technology (such as social media) that contains eating disorder–related content.

We focus on weight intentions and users' perceptions of the impact of a weight loss app (DropPounds) on disordered eating behaviors. This study adds to research on weight loss apps and communities, as well as to our understanding of the use of technology by those with disordered eating behaviors and eating disorders.

Two primary research questions (RQ) guided this study, which are as follows:

- RQ1: What is the composition of the overall app community in terms of body mass index (BMI) goals?
- RQ2: What are users with underweight BMI goals perceptions of the app in relation to disordered eating behaviors?

To answer these RQs, we first analyzed users' profile data to get a better understanding of the composition of the community. Then we provided examples of how users with underweight BMI goals perceive the app. This exploratory study is important for our understanding of the different types of users who utilize weight loss apps and how weight loss apps may impact users with a desire to be underweight. However, more questions emerged, and therefore, future research is needed.

To our knowledge, this study is one of the first to (1) provide an analysis of users' BMIs in an app and community geared toward weight loss to highlight potential disordered eating behaviors and (2) to consider how users with a desire to be underweight understand weight loss apps in relation to disordered eating behaviors. In this paper, we first explain why we used underweight BMI goals as a marker for potentially disordered eating behaviors. Then we explain our methods and findings. Finally, we discuss what can be learned from this study and future work.

Desire to Be Underweight and Disordered Eating Behaviors

Using weight loss apps while wanting to be underweight presents a number of issues related to eating disorders and disordered eating behaviors, which is why we focused on underweight BMI goals and the use of weight loss apps to achieve those goals. Whereas anorexia nervosa is partially characterized by being underweight [19] and a refusal to be a healthy weight (as cited in [20]), intentions to be underweight for one's height itself are problematic. To determine which users wanted to be underweight, we used BMI. For adults 20 years or older, an underweight BMI is less than 18.5 kg/m^2.

Setting unhealthily low BMI goals signals a desire to be underweight and a drive for thinness, which are associated with disordered eating behaviors and eating disorders [21,22]. Drive for thinness "is characteristic of individuals with fear of weight gain who diet to prevent it, but also of those who seek to attain an unhealthily low body weight as seen in many individuals with anorexia nervosa or bulimia nervosa" [22]. Pro-eating disorder blogs and websites are often characterized by users' drive to be underweight [23-25]. In terms of goal setting, Boero and Pascoe [21] found that many users of pro-ana (pro-anorexia nervosa) communities write about their maximum, current, and goal weights and are self-identified or diagnosed with eating disorders. The reported goal weights are to be underweight even if users' current weights are in the healthy or underweight range [21].

The desire to be underweight coupled with the use of weight loss apps may put users at risk for or signal disordered eating behaviors. Weight loss apps encourage dieting behaviors to reach those who need to lose weight. Unfortunately, dieting behaviors are linked to the development of eating disorders [26-28]. In fact, women who severely diet are 18 times more likely to develop an eating disorder, and those who moderately diet are 5 times more likely to develop an eating disorder than those who do not [27]. Studies of adolescent girls have found that high BMI is not a factor for dieting initiation; many girls

of a healthy weight and even those who are underweight report wanting to lose weight and to go on a diet [29,30]. Simply wanting to lose weight is associated with disordered eating and weight control behaviors. In her study of US high school students, Forman-Hoffman [31] found that dieting and exercising to lose weight were linked to more unhealthy eating and weight control behaviors [31]. Additionally, over one-third of students who wanted to lose weight also reported one or more of those disordered behaviors [31]. Thus, this drive to be underweight acted as a marker for potential disordered eating behaviors.

Methods

Approach and Analysis

We had profile data and forum posts from DropPounds, a mobile- and Web-based weight loss app available on iPhone operating system (OS, Apple Inc.), Android, and through the Web. This app was chosen because at the time the data were provided, it was a popular health app and includes many of the features and content found in the majority of weight loss apps today. For example, DropPounds allows users to track their diet and physical activity. It also has an optional online community associated with it, which users can turn to for advice and support.

We present two pieces of this study that move from a broader picture of the DropPounds community to a narrower view of specific DropPounds users. First, we analyzed users' profile data to get a sense of the number of users within the community who set underweight goals. From the forum data, we provided examples of users' posts to highlight their perceptions of the app's effect on disordered eating behaviors.

Due to the sensitivity around eating disorders, it was important to consider the ethics around conducting research in an online forum with users who may have a history with eating disorders [32]. Both the name of the app and community were changed to protect users' privacy. Also, we changed the wording of quotations slightly to ensure the anonymity of users while still maintaining the tone and theme of the quotation. As the app company provided the data, we did not have access to users' information to contact them for their consent.

Institutional review board approval was obtained from 3 universities to conduct the research. The company that owns, maintains, and operates the app gave permission to conduct research and provided app data, including forum posts and portions of users' profile data. Before providing the data, the company assigned a random unique identifier to users. They did not provide fields that were individually identifying. Forum data and parts of profile data are publicly accessible through the app. Anyone can create an account for free and read posts on the forums. Users input their height, current weight, and then set a goal weight and how many pounds per week they want to lose (up to 2 pounds).

Users' Profile Data

DropPounds provided us with profile data from 19,710 users in 2012. To analyze the overall community, we removed users if they had a BMI under 5 or over 125 (n=14), did not have weight or height data available (n=271), were younger than 20 years (because of the way BMI is calculated for those under 20 years; n=832), or over 99 years (n=1). That left us with 18,601 users. We calculated users' BMI from the weight and height data in their profile. For adults ≥20 years, underweight is <18.5, healthy weight is 18.5 to 24.9, overweight is 25.0 to 29.9, and obese is 30+. Looking at profile data gave us a better understanding of the composition of the overall DropPounds community.

For the users' profile data, we used Excel (Microsoft Corp.) to investigate the types and frequency of users' BMI at the time they created their weight loss plan (start BMI), BMI at the time of data collection (current BMI), and goal BMI.

Users' Posts

We analyzed content from the DropPounds online community for two reasons. First, we wanted to observe whether weight loss app users discussed not only the role of the app but also how it impacted disordered eating behaviors. The examination of forum posts revealed that this phenomenon was occurring. Second, we wanted users' perceptions without the influence of a researcher asking specific questions. The dataset used in this study comprised 321,999 posts over 24,183 threads that were created from October 2009 to July 2012. Whereas the design of specific features has changed since 2012, the types of features have remained consistent.

To isolate discussions about eating disorders, we used a number of eating disorder–related keywords to identify candidate threads. Keywords included an[eo]rexi[ac], ana, bul[ie]mi[ac], mia, compulsive overeating, body d[iy]smorphi[ac] disorder, bing[e], eating disorder, ED, purg[e], and EDNOS (eating disorder not otherwise specified). The initial keyword search returned 6190 threads representing 9255 unique users. The first author examined the initial set of threads to identify and remove any irrelevant content that the keyword search returned. After removing irrelevant threads, we identified 1036 relevant threads (2678 posts). After removing duplicates (n=342), we had 2336 posts that represented 1080 unique users.

We then pulled every post in the dataset written by users with underweight BMI goals (n=246). After becoming familiar with the data, the first author wrote notes about each post, including the content and disordered eating behaviors mentioned and then created codes based on the content (eg, app impact, binge triggers, and community support). Coded posts were then grouped together. We chose to focus on app impact. We reviewed these posts to see how users discussed the impact of the app on disordered eating behaviors and grouped these posts into two broad categories: (1) reduces disordered eating behaviors and (2) exacerbates disordered eating behaviors. We then broke these up into smaller groups to highlight specific ways the users believe the app reduces or exacerbates disordered eating behaviors, with the purpose of providing example quotations about the app's impact on disordered eating behaviors for other researchers to use as a basis for future research.

Results

In this section, we first present statistics about users' profile data related to their BMI and goals to show the prevalence of users with underweight BMI goals utilizing the app. Then we present examples of how these users discussed the impact of the app on disordered eating behaviors.

Users' Profile Data (RQ1)

Of 18,601 users, 14,031 identified as female and 4570 as male. The reported age of users ranged from 20 to 99 years (mean=38.42, standard deviation [SD]=11.71, median=37, mode=28). As shown in Figure 1, the majority of users start with weight loss goals (n=18,370), followed by maintenance goals (n=231). None of the users start with weight gain goals. Looking at the change between users' current BMI to goal BMI, we show most users also have weight loss goals (n=17,984), followed by weight gain goals (n=311) and maintenance goals (n=306). In some instances, users lose more weight than they planned from the time between their program start weight and current weight, so their current BMI to goal BMI reflects a weight gain goal even though their overall objective is to lose weight.

Within the community, 2.18% (406/18,601) start with underweight BMIs, 2.81% (522/18,601) are currently underweight, and 6.78% (1261/18,601) of the community have a desire to be underweight, which can be seen in Figure 2 along with other information about the start, current, and goal BMIs of the community. The majority of users with underweight BMI goals identify as female (n=1238 compared with n=23 who identify as male). Of those users with underweight BMI goals, most want a BMI of 17 or above (n=585), followed by under 15 (n=289), 16 to 16.99 (n=220), and 15 to 15.99 (n=167).

Of the users with underweight BMI goals, the majority (n=671) had healthy start BMIs, followed by underweight start BMIs (n=406), then overweight start BMIs (n=125), and finally obese start BMIs (n=59), as shown in Figure 3. All users with underweight start BMIs (n=406) had underweight goal BMIs; that is, no one who was underweight when they began the program wanted to gain weight to be in the healthy range. Additionally, none of the users with underweight start BMIs had healthy, overweight, or obese current BMIs.

Most users with underweight goal BMIs wanted to lose weight (n=1237), and a small subset wanted to maintain their weight (n=24). None of the users with underweight BMI goals wanted to gain weight. Thus, none were interested in gaining weight even if that allowed them to remain underweight.

Figure 1. Number of users who had weight loss goals and weight gain goals.

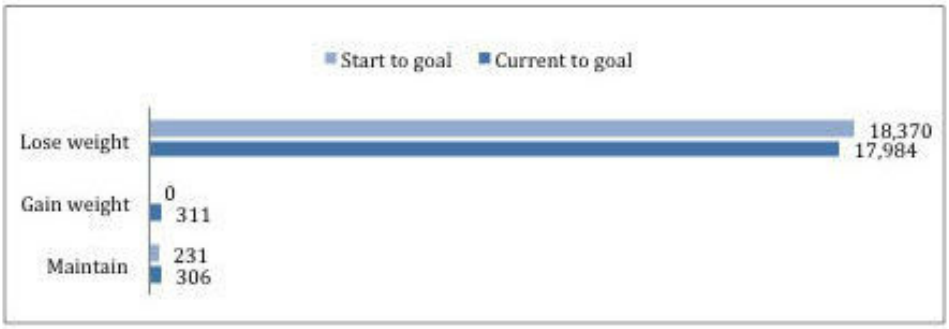


Figure 2. Number of users whose body mass index (BMI) was underweight, healthy weight, overweight, and obese at start, current, and goal.

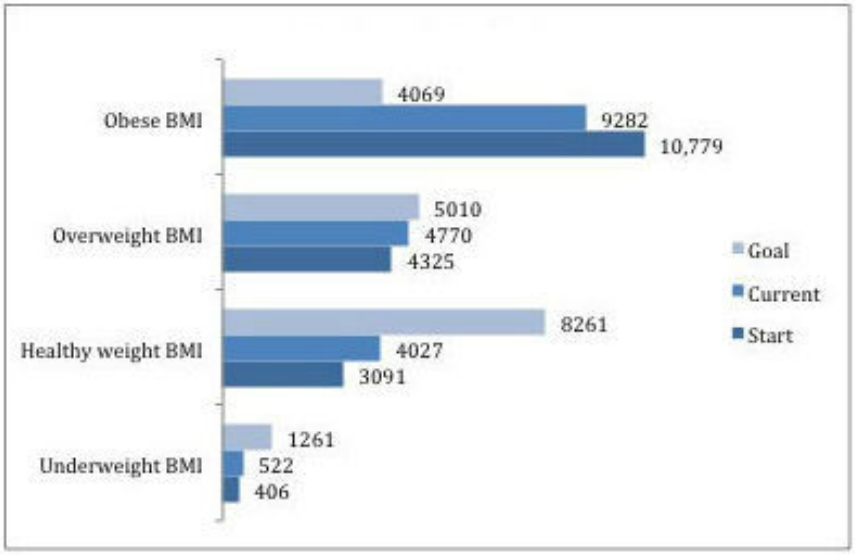
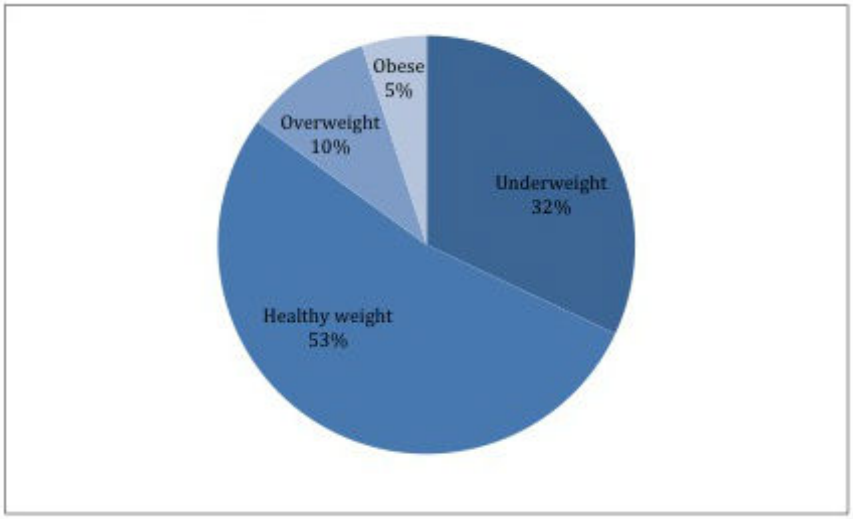


Figure 3. Number of users with underweight body mass index (BMI) goals who have underweight, healthy weight, overweight, and obese start BMIs.



Users' Posts (RQ2)

We found that 8.98% (97/1080) of users who post in the forums about eating disorders have underweight BMI goals. We then looked at all users with underweight BMI goals and found 7.69% (97/1261) of them post in the forums about eating disorders. Users with underweight BMI goals produced a total of 246 posts (mean=2.54, SD=2.29, range 1-24). Thus, 10.53% (246/2336) of posts about eating disorders were written by users with underweight BMI goals. Seventeen posts from 13 users contained content related to how the app affects disordered eating behaviors. Details on these 13 users are provided in Table 1.

Table 1. Users' age, body mass index (BMI), symptoms, and app perception.

User ID	Reported age, in years	Disordered eating behaviors	Start BMI[a]	Current BMI	Goal BMI	Exacerbates	Reduces
62663	34	Restriction, low weight	12.55	13.11	12.55		X
92214	27	Control food intake	16.09	16.49	14.51	X	
23774	24	Bingeing	17.22	17.22	15.78		X
144279	52	Restriction, bingeing	17.28	15.06	14.62		X
154295	22	EDNOS[b], restriction, bingeing	17.80	17.47	14.98		X
274788	40	Bingeing	18.16	16.09	15.21		X
533842	40	Anorexia and bulimia nervosa	18.26	15.70	15.21	X	
2042	48	Bingeing	18.46	18.46	16.98		X
29213	20	Restriction, purging	18.75	18.46	16.24	X	
318229	23	Anorexia and bulimia nervosa	19.57	18.47	18.23	X	
172884	41	Former bulimia nervosa	20.90	20.01	18.07	X	X
215596	54	Bingeing, emotional eating	21.43	18.04	18.04		X
397752	24	Bingeing	22.78	20.88	18.31		X

[a]BMI: body mass index.

[b]EDNOS: eating disorder not otherwise specified.

All users with underweight BMI goals who posted about the effects of the app in the forum identified as female. The majority of these users began the program underweight (n=8), followed by healthy weight (n=5). For current BMI, most users fell into the underweight category (n=11), followed by healthy weight (n=2). Eight users discussed how the app helps reduce disordered eating behaviors, 4 users discussed how the app exacerbates disordered eating behaviors, and 1 user talked about how the app did both. We provide example posts that highlight these perceptions of the users about the app in Tables 2 and 3.

More often than not, users felt that DropPounds was a positive influence because it created awareness and accountability that reduced bingeing, helped them eat more, improved their food choices, and provided them with a healthy plan. Many posts focus on bingeing behaviors, so many users felt the app helped them control those behaviors and choose healthier foods overall.

Users with a history of extreme calorie and food restriction felt that the app gave them awareness about their restrictive behaviors, which helped them see where they should add foods. Some users believed the daily calorie budget was inherently healthy. Table 2 provides example posts from users who discussed these positive aspects of the app.

Table 2. Positive aspects of the app and example posts.

Positive aspects	Example post
Reduces bingeing	"The best thing for me for emotional eating or binge eating is logging! If I am faithfully logging, I have much better control over that stuff because I don't want to enter a bunch of crap. I am really proud of myself for not succumbing to those desires to binge." [ID 215596]
Helps eat more	"I suffered from disordered eating my entire life. My eating issues were never extreme enough to be considered full-blown eating disorders but were enough to have a big negative effect on my life. Until I started using DropPounds, it was almost impossible for me to eat 3 meals a day. My usual pattern included starving myself and then bingeing, compulsive eating, and sporadic, unsustainable diets. My self-esteem has always been tied to my weight and whether I had a 'good eating day' or a 'bad eating day.' Since I have been on DropPounds, I have finally learned how to eat 3 meals a day (and snacks). Every meal and mouthful is still a battle but at least I'm finally winning the fight." [ID 144279]
Improves food choices	"DropPounds for me is more about being held accountable for my food choices, as I have a bit of a sugar issue and tendency to binge until I feel ill. This is about making sure I get enough fruits, vegetables, fiber , and avoid eating mindlessly." [ID 23774]
Provides a healthy plan	"I have EDNOS, and I'm trying to recover. This app and community really motivate me to lose weight in a healthy manner. Unfortunately, I purged last night, but today I didn't. At one point, I used to purge everything I ate no matter what it was: fruit, vegetables, diet coke, and water. I'm motivated and inspired to finish recovery by myself because I had a negative experience in a hospital. I'm very happy here, and I love how it [DropPounds] gives you the amount of calories to eat. You can still lose weight, and it selects a goal for you, which makes it healthy." [ID 154295]

Some users also discussed how the app could exacerbate disordered eating behaviors. For example, the app encourages purging calories through excessive exercise by providing negative feedback when users exceed their budget and allowing them to erase calories to receive positive feedback. Not only does it promote compensatory behaviors, but it also encourages users to eat less than their allotted budget. Having disordered eating behaviors in combination with using the app also could lead to or exacerbate obsessive behaviors and thoughts around logging and numbers. Whereas some users felt the app's algorithm automatically meant that the plan was healthy, other users suggested that the app's goal-based plan was actually unhealthy. Table 3 provides example posts of these negative aspects. These findings provide a preliminary look into how apps may reduce and exacerbate disordered eating behaviors.

Table 3. Negative aspects of the app and example posts.

Negative aspects	Example post
Encourages purging	"The times where my bar showed I was over calories, I would punish myself with an extensive amount of exercise while talking down to myself. There was a times where I would go over an insignificant amount of calories, 50 perhaps, and punish myself with a large amount of unnecessary exercising." [ID 29213]
Promotes eating less	"I have found myself doing this [trying to eat less and less like it's a game] and have to remind myself daily (usually at every meal) that it's not about the number; it's about making healthy choices . I have to make myself not feel like an utter failure if I don't stay under my calorie limit." [ID 533842]
Leads to obsessive behaviors	"I have struggled with bulimia/anorexia for the past 4 to 5 years, and I still struggle today in being happy with my body. I am a perfectionist and have a somewhat obsessive personality so I can get obsessed with logging my food and thinking about how many calories I am eating and drinking at all times. I probably shouldn't be on this site sometimes!" [ID 318229]
Provides a dangerous plan	"Many people who frequent these forums know about the 1200/1500 calorie minimum for women and men, but those who have not joined the forums only assume that the less they eat, the more they will lose. For example, when I began this program over a year ago, I set my goal to 2 pounds per week in order to get things accomplished faster. My budget was around 900 calories, which I ate. DropPounds is the one calculating the calories people consume. While we cannot solely blame DropPounds for its cold calculation, we have to consider the ignorance of many people who are using this program and who are destroying their well-being in the process." [ID 29213]

Discussion

Principal Findings

In summary, we found three main results: (1) no user (regardless of start BMI) starts with a weight gain goal, and most users want to lose weight; (2) 6.78% (1261/18,601) of the community want to be underweight, and most identify as female; (3) users with underweight BMI goals tend to view the app as positive; however, some acknowledge its role in exacerbating disordered eating behaviors. In this section, we discuss these findings and present a number of areas that researchers and designers need to consider in more detail.

Weight Loss and Underweight Goals

We found that no user set a weight gain goal. In fact, the vast majority of the community (98.76%; 18,370/18,601) begins the program to lose weight (not maintain or gain). The users who

are underweight when they begin the program do not want to gain weight, and most of these users want to lose additional weight, which would put them at a more extreme low weight for their height. A small subset of users with underweight BMI goals is using the app to maintain an already low weight. No one using the app is doing so to gain weight even if they should gain weight or report needing to gain weight as part of their eating disorder recovery. Thus, the overall focus of app use is weight loss irrespective of the start weight of users.

In addition to the heavy focus of weight loss by all users, we found that 6.78% (1261/18,601) of the users wish to be underweight according to BMI. The majority of users who set underweight goals begin the program at either a healthy weight or are already underweight, and most of these users identify as female. These findings suggest women often want to lose weight even when weight loss is unnecessary, which is in line with prior research [31,33]. This may be explained in part by women's body dissatisfaction and weight perception. It is common for women to be discontent with their bodies and weight. In fact, in North America, there is such a pervasive body dissatisfaction and preoccupation with weight among women that psychologists have developed a term "normative discontent," which describes the normalcy of being unhappy with one's weight as a woman [34].

Research has also shown that women and girls tend to have an inaccurate perception of their weight; they perceive their weight as higher than it actually is [30,31,35]. For example, in their study of first-year college women, Cilliers et al [30] found that only a few in the healthy weight range were satisfied with their weight, and many wanted to lose weight, which could result in their engaging in unnecessary and unhealthy weight control practices. A quarter of the underweight students also still wanted to lose weight, and the remaining underweight students wanted to keep their underweight status [30]. Similarly, Forman-Hoffman [31] found that over 20% of high school students overestimate or extremely overestimate their weight. When women and girls perceive their weight as higher than it is, they may wish to lose weight that would put them at an unhealthy weight for their height. This overestimation is linked to disordered eating behaviors [31]. Thus, users' goals may explain a lot about them. To promote health, researchers should further examine the relationship between perceived weight, goal weight, and eating disorders to find ways to empower users to make healthy choices and set healthy goals.

Although DropPounds and other weight loss apps and communities are designed for users who need to lose weight, our study shows that users who likely do not need to lose weight but are dissatisfied with their current weight are using the app to achieve unhealthily low weights. This is an important finding because designers and developers often focus on the intended and "ideal" user, which makes sense, given the focus of the app (weight loss); however, our study shows that we need to pay more attention to unintended or "nonideal" users and the uses, perceptions, and effects of weight loss apps on them.

App Perceptions

We provided example posts from users who discussed the app as positive and negative. These examples are not meant to be an exhaustive list of the use and perceptions of weight loss apps; instead the intention is that these provide a basis for researchers to thoroughly investigate the role of weight loss apps for those with disordered eating behaviors. Researchers can use these eight themes (reduces bingeing, helps eat more, improves food choices, provides a healthy plan, encourages purging, promotes eating less, leads to obsessive behaviors, and provides a dangerous plan) to examine aspects and features of weight loss apps that could be useful for eating disorder recovery or aggravate disordered eating behaviors.

Similar to Tan et al [15], we found that users discussed the weight loss app as reducing disordered eating behaviors and exacerbating them; however, more users talked about the positive aspects of the app than the negative aspects. These findings are both in line with and in opposition to interview-based research about the use and perceptions of weight loss apps by women with eating disorders. Interview-based findings support the idea that weight loss apps encourage purging or compensatory behaviors, promote restriction, and lead to obsessive logging [14]. However, this study did not discuss dangerously low plans, as the focus was on a different app that did not allow users to set a daily calorie budget below 1200 [14], suggesting that having minimum daily calorie budgets may be beneficial to users, especially those who are unaware of how many calories they need. Additionally, positive aspects of the app included its ability to show users how much they needed to eat, much like this study; however, the set healthy plan did not emerge as a major finding [14].

So why do users in this study tend to view the app as mostly positive? One possible explanation may be related to the type of disordered eating behaviors. Of the 13 users, 10 mentioned bingeing or bulimia-related behaviors. The app may, in fact, be beneficial to promote mindfulness, and logging foods during a binge may reduce the amount of food users eat. However, given that these users have a desire to be underweight, the app may contribute to or exacerbate other behaviors. Thus, more research is needed to differentiate the effects and uses related to types of disordered eating behaviors and eating disorders. For instance, creating awareness may be beneficial for those with binge eating disorder, but this awareness could be problematic for those who have a high drive for thinness, fear of weight gain, anorexia nervosa, and so on.

Another possible explanation may relate to users' current stage of their disordered eating or eating disorder and their ability to reflect on their behaviors. When users post in the forums about the app, they may feel as though the app is helpful to them even if their behaviors are disordered. They may be at a stage where they do not recognize their disordered eating behaviors or are unwilling to. This is echoed in the concept of a user's health journey [14]. In the interview-based work, participants were reflecting on their use over time, and they often explained that at the beginning of app use they saw no problem with their behaviors or the app but later realized the app was exacerbating their disordered behaviors more than helping them [14]. Thus, it is possible that the users in this study were in the earlier stages of use when they discussed the app in the forums. More research is needed to understand the effects of weight loss apps on

disordered eating behaviors and how the stages of the users' health journey change the role technology plays.

Discrepancies Between App Perceptions and Goals

On the basis of this study alone, we do not understand why users set low weight goals. Some users explicitly stated needing to gain weight for recovery but had weight *loss* goals. For example, users with underweight start BMIs and underweight goal BMIs said that the technology is helping to reduce their eating disorder behaviors, and every user with an underweight start weight had an underweight goal weight. This begs the question: do some users really want to recover? We cannot determine whether or not users truly want to use the app to reduce their disordered eating behaviors. Our results indicate that women's discussions of eating behaviors in the DropPounds forum often focused on being healthy, and no user in the sample openly said they were trying to maintain disordered eating behaviors. However, their private weight goals would put them at an underweight BMI.

The tension between users' private goals and forum posts may be related to self-presentation. According to Counts and Stecher [36], "self-presentation can be thought of as the image or idea of the self, or the process of creating this image for a variety of social purposes." Research has shown that people present themselves in desirable ways when they are in the public [37]. Self-presentation motives include achieving your goals, presenting a positive view of self to the world, and conforming to social norms. According to Goffman [37], life is comparative to the theater: people do "front stage" work when they are interacting with others in public settings, and they do "back stage" work, which comprises the private things people do when no one is looking. These two presentations can be misaligned. In this case, the forum posts represent "front stage" work, whereas the private profile goals represent "back stage" work.

Although research has shown that eating lightly to achieve thinness is desirable for women [38], having an eating disorder or disordered eating behaviors in a community whose focus is health may be undesirable. This is in line with prior work on self-presentation in an online community. Schwammlein and Wodzicki [39] found that "members of the common-identity community focused on characteristics shared among members of the community" versus those who were part of the common-bond community. The DropPounds community shares a common identity: its users emphasize *healthy* weight loss. Many of its users are serious about maintaining a healthy lifestyle and are against unhealthy or extreme tactics to lose weight. Due to this, users who may have disordered eating behaviors and want to post in the forums may attempt to conform to the community's norms and its members' opinions and characteristics. Many more may choose not to post in the forums about their disordered eating behaviors because of fear of backlash from the rest of the community or privacy concerns. Research has shown that women with eating disorders may be reluctant to use social and community features of weight loss apps [40]. This may explain why only 97 users with underweight BMI goals were present in the forums even though 1261 users set underweight BMI goals and why only 13 of these users talked about the impact of the app within the forums.

We cannot determine whether users are actively trying to maintain their disordered eating behaviors or whether they want to recover but are not setting appropriate goals. There may be a subset of users who do not want to recover from their eating disorder or disordered eating behaviors. Weight loss apps and forums can give them a false feeling that their behaviors are healthy, which can allow them to deny they have disordered eating behaviors. Researchers need to investigate in more detail why users with disordered eating behaviors use weight loss apps and their intentions regarding eating disorder maintenance and recovery.

Another potential explanation comes out of the interview-based study [14]. In that study, users discussed trying to use the app for recovery but falling back into old habits [14]. Therefore, some users may in fact be trying to recover while combatting the urge to lose weight and restrict calories. This may, in part, be because of the nature of weight loss apps. Through design, users are supposed to be motivated to lose weight, so even when they set weight gain goals or attempt to eat more calories, the app still shows visualizations to motivate them to lose weight. Thus, more research is needed on the effects of design and how we motivate users through design. Health app designers should also consider ways to promote other types of goals besides weight loss. Weight loss is not the only type of health and fitness goal. For instance, users may wish to gain muscle or use the app to achieve exercise-related goals not even tied to a goal weight.

Limitations

There are a few limitations of this research, including sample size, date of data collection, and using BMI. Although there were over 1200 users who had underweight BMI goals, not every user posted in the forums. Analyzing the profile and forum data meant that we were unable to get the perspectives of users with disordered eating behaviors who did not post in the forum. Additionally, for the purposes of this study, we focused on users' discussions on the impact of the app. Thus, we could only provide example quotations from a small subset of users. This likely does not cover all perceptions of the app but is meant as a jumping point for future research.

Another limitation is the age of the data. Whereas certain features of the app have changed since the time of data collection, the overall focus of the app is the same. The app still contains a food, exercise, and weight loss log and shows progress visualizations based on these factors, but the look of the app has changed. However, the focus of this research was not on specific design features or aesthetics, so the findings are still relevant not only to DropPounds but also to other weight loss apps. Since data collection, the app's popularity has skyrocketed. Thus, we suspect even more users with disordered eating behaviors are utilizing the app.

Although BMI has its limitations, BMI is used in the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-V) to aid in the diagnosis of eating disorders [19]. Additionally, per the Centers for Disease Control and Prevention (CDC), BMI below 18.5 is considered underweight [41]. We chose to focus on users with underweight BMI goals because this desire to be underweight can signal unhealthy behaviors.

Having an underweight BMI goal does not mean that all of these users have clinical eating disorders; rather the intent to be underweight and wanting to lose weight when already at a healthy weight or underweight could be indicative of disordered eating behaviors [21,22]. Additionally, BMI under 18.5 is a potential marker of anorexia nervosa. Thus, we used goal BMI as an indicator of disordered eating behaviors. However, it is likely that we missed users with disordered eating behaviors who have goal weights that put their BMI in the healthy, overweight, or obese range, which means that there are potentially many more users with disordered eating behaviors using weight loss apps. The aim of this study was to focus on weight intentions, specifically users with underweight BMI goals, and shed light on their app perceptions. Despite the limitations, this study is a good first step toward looking at the composition of weight loss app communities and the impact of weight loss apps on disordered eating behaviors.

Future Work

This research represents one phase of a larger project. We are also conducting interviews with weight loss app users with eating disorders. From the profile data, we found that many users had either underweight BMI start weights or underweight BMI goals, which suggests a large prevalence of users with disordered eating behaviors. However, most of these users did not post about eating disorders in the forums. Thus, conducting interviews allows us to get a larger sample size as well as ask additional questions about how users utilize weight loss apps and how those apps impact them. After we conduct the interviews, we plan to get a broader view of the phenomenon by conducting a survey, which will help with generalizability.

Conclusions

In this study, we looked at the underweight BMI goals of users in a weight loss community and perceptions of these users with regard to the impact of a weight loss app on disordered eating behaviors. A number of users within the community had underweight BMI goals, suggesting they have a strong drive for thinness and a desire to be underweight, which may signal disordered eating behaviors. Users with underweight BMI goals tend to view the app as beneficial for disordered eating behaviors, especially bingeing. Although users with underweight BMI goals felt different features of the app could both reduce and exacerbate disordered eating behaviors, their overall perceptions and goals were misaligned. While this study provides examples of how weight loss apps may impact disordered eating behaviors, a number of questions emerged leading to suggestions for additional research directions. As this is an understudied area, more work is needed on the use of weight loss apps by users with eating disorders and disordered eating behaviors. Our future work aims to explore this area more thoroughly. We hope that this study sparks more research on the role of technology for users with disordered eating behaviors.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DGE1255832. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Conflicts of Interest

None declared.

References

1. Wade T, Keski-Rahkonen A, Hudson J. Epidemiology of eating disorders. In: Textbook in Psychiatric Epidemiology. 3rd edition. New York: Wiley; 2011:343-360.
2. Reba-Harrelson L, Von HA, Hamer RM, Swann R, Reyes ML, Bulik CM. Patterns and prevalence of disordered eating and weight control behaviors in women ages 25-45. Eat Weight Disord 2009 Dec;14(4):e190-e198 [FREE Full text] [Medline: 20179405]
3. Nagl M, Jacobi C, Paul M, Beesdo-Baum K, Höfler M, Lieb R, et al. Prevalence, incidence, and natural course of anorexia and bulimia nervosa among adolescents and young adults. Eur Child Adolesc Psychiatry 2016 Aug;25(8):903-918. [doi: 10.1007/s00787-015-0808-z] [Medline: 26754944]
4. Ransom DC, La GJG, Woody EZ, Boyd JL. Interpersonal interactions on online forums addressing eating concerns. Int J Eat Disord 2010 Mar;43(2):161-170. [doi: 10.1002/eat.20629] [Medline: 19308991]
5. Keski-Rahkonen A, Tozzi F. The process of recovery in eating disorder sufferers' own words: an internet-based study. Int J Eat Disord 2005;37 Suppl:S80-6; discussion S87. [doi: 10.1002/eat.20123] [Medline: 15852327]
6. Bowler L, Oh J, He D, Mattern E, Jeng W. Eating disorder questions in yahoo! answers: information, conversation, or reflection? In: Proc Am Soc Info Sci Tech. 2012 Presented at: ASIST; October 28-31, 2012; Baltimore, MD p. 1-11. [doi: 10.1002/meet.14504901052]
7. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
8. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]

9. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
10. 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]
11. Rooksby J, Rost M, Morrison A, Chalmers M. Pass the ball: enforced turn-taking in activity tracking. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, New York: ACM; 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, Republic of Korea p. 2417-2426. [doi: 10.1145/2702123.2702577]
12. Cordeiro F, Bales E, Cherry E, Fogarty J. Rethinking the Mobile Food Journal: Exploring Opportunities for Lightweight Photo-Based Capture. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. New York, New York: ACM; 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, Republic of Korea p. 3207-3216. [doi: 10.1145/2702123.2702154]
13. Choe E, Lee B, Munson S, Pratt W, Kientz J. Persuasive performance feedback: the effect of framing on self-efficacy. AMIA Annu Symp Proc 2013;2013:825-833 [FREE Full text] [Medline: 24551378]
14. Eikey EV. Proceedings of the 2017 CHI conference on human factors in computing systems. New York, New York: ACM; 2017 Presented at: CHI '17; May 6-11, 2017; Denver, Colorado p. 642-654. [doi: 10.1145/3025453.3025591]
15. Tan T, Kuek A, Goh SE, Lee EL, Kwok V. Internet and smartphone application usage in eating disorders: a descriptive study in Singapore. Asian J Psychiatr 2016 Feb;19:50-55. [doi: 10.1016/j.ajp.2015.11.007] [Medline: 26957338]
16. Peebles R, Wilson JL, Litt IF, Hardy KK, Lock JD, Mann JR, et al. Disordered eating in a digital age: eating behaviors, health, and quality of life in users of websites with pro-eating disorder content. J Med Internet Res 2012 Oct 25;14(5):e148 [FREE Full text] [doi: 10.2196/jmir.2023] [Medline: 23099628]
17. Chancellor S, Pater J, Clear T, Gilbert E, De Choudhury M. #thyghgapp: instagram content moderation and lexical variation in pro-eating disorder communities. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, New York: ACM; 2016 Presented at: CSCW '16; February 27-March 2, 2016; San Francisco, CA p. 1201-1213. [doi: 10.1145/2818048.2819963]
18. Pater J, Haimson O, Andalibi N, Mynatt E. "Hunger Hurts but Starving Works": Characterizing the presentation of eating disorders online. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. New York, New York: ACM; 2016 Presented at: CSCW '16; February 26-March 2, 2016; San Francisco, CA p. 1185-1200. [doi: 10.1145/2818048.2820030]
19. Diagnostic Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC: American Psychiatric Association; 2013.
20. Le Grange D, Doyle PM, Swanson SA, Ludwig K, Glunz C, Kreipe RE. Calculation of expected body weight in adolescents with eating disorders. Pediatrics 2012 Jan 04;129(2):e438-e446 [FREE Full text] [doi: 10.1542/peds.2011-1676] [Medline: 22218841]
21. Boero N, Pascoe C. Pro-anorexia communities and online interaction: bringing the pro-ana body online. Body & Society 2012 May 24;18(2):27-57. [doi: 10.1177/1357034X12440827]
22. Peñas-Lledó E, Bulik CM, Lichtenstein P, Larsson H, Baker JH. Risk for self-reported anorexia or bulimia nervosa based on drive for thinness and negative affect clusters/dimensions during adolescence: A three-year prospective study of the TChAD cohort. Int J Eat Disord 2015 Sep;48(6):692-699 [FREE Full text] [doi: 10.1002/eat.22431] [Medline: 26013185]
23. Gies J, Martino S. Uncovering ED: a qualitative analysis of personal blogs managed by individuals with eating disorders. The Qualitative Report 2014 Jul 21;19(29):1-15 [FREE Full text]
24. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: a content analysis of pro-eating disorder web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]
25. Coker E, Abraham S. Body weight dissatisfaction: a comparison of women with and without eating disorders. Eat Behav 2014 Aug;15(3):453-459. [doi: 10.1016/j.eatbeh.2014.06.014] [Medline: 25064299]
26. Shisslak CM, Crago M, Estes LS. The spectrum of eating disturbances. Int J Eat Disord 1995 Nov;18(3):209-219. [Medline: 8556017]
27. Neumark-Sztainer D, Wall M, Guo J, Story M, Haines J, Eisenberg M. Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: how do dieters fare 5 years later? J Am Diet Assoc 2006 Apr;106(4):559-568. [doi: 10.1016/j.jada.2006.01.003] [Medline: 16567152]
28. Ackard DM, Croll JK, Kearney-Cooke A. Dieting frequency among college females: association with disordered eating, body image, and related psychological problems. J Psychosom Res 2002 Mar;52(3):129-136. [Medline: 11897231]
29. Huon G, Lim J. The emergence of dieting among female adolescents: age, body mass index, and seasonal effects. Int J Eat Disord 2000 Sep;28(2):221-225. [Medline: 10897085]
30. Cilliers J, Senekal M, Kunneke E. The association between the body mass index of first-year female university students and their weight-related perceptions and practices, psychological health, physical activity and other physical health indicators. Public Health Nutr 2006 Apr;9(2):234-243. [Medline: 16571178]

31. Forman-Hoffman V. High prevalence of abnormal eating and weight control practices among U.S. high-school students. Eat Behav 2004 Nov;5(4):325-336. [doi: 10.1016/j.eatbeh.2004.04.003] [Medline: 15488447]
32. Stommel W. Entering an Online Support Group on Eating Disorders: A Discourse Analysis. Amsterdam/New York, NY: Rodopi BV; 2009:1-286.
33. Lynn AL. UNL. 2012. Body mass index trends and nutrition goals of college students between 2007-2011 URL: http://digitalcommons.unl.edu/nutritiondiss/38/ [accessed 2017-09-19] [WebCite Cache ID 6taiV3gyB]
34. Spettigue W, Henderson KA. Eating disorders and the role of the media. Can Child Adolesc Psychiatr Rev 2004 Feb;13(1):16-19 [FREE Full text] [Medline: 19030149]
35. Martin MA, Frisco ML, May AL. Gender and race/ethnic differences in inaccurate weight perceptions among U.S. adolescents. Women's Health Issues 2009;19(5):292-299 [FREE Full text] [doi: 10.1016/j.whi.2009.05.003] [Medline: 19733799]
36. Counts S, Stecher K. Self-presentation of personality during online profile creation. In: Proceedings of the Third International Conference on Weblogs and Social Media. 2009 Presented at: ICWSM (International Conference on Weblogs and Social Media); May 17-20, 2009; San Jose, CA.
37. Goffman E. The Presentation of Self In Everyday Life. Garden City, NY: Doubleday; 1959.
38. Vartanian LR, Herman CP, Polivy J. Consumption stereotypes and impression management: How you are what you eat. Appetite 2007 May;48(3):265-277. [doi: 10.1016/j.appet.2006.10.008] [Medline: 17157957]
39. Schwämmlein E, Wodzicki K. What to tell about me? self-presentation in online communities. J Comput Commun 2012;17(4):387-407. [doi: 10.1111/j.1083-6101.2012.01582.x]
40. Eikey EV. Privacy and weight loss apps: a first look at how women with eating disorders use social features. In: Proceedings of the ACM 2016 International Conference on Supporting Group Work.: ACM; 2016 Presented at: GROUP '16; November 13-16, 2016; Sanibel Island, FL. [doi: 10.1145/2957276.2996282]
41. CDC. 2015. About adult BMI URL: http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/ [accessed 2017-09-08] [WebCite Cache ID 6tKQKF5B1]

Abbreviations

BMI: body mass index
CDC: Centers for Disease Control and Prevention
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, 5th edition
EDNOS: eating disorder not otherwise specified
mHealth: mobile health
OS: operating system
RQ: research question
SD: standard deviation

Edited by G Eysenbach; submitted 22.09.16; peer-reviewed by D Kolar, S Muir, R von Brachel, E van Furth, L Santos-Merx; comments to author 08.01.17; revised version received 03.04.17; accepted 12.07.17; published 12.10.17.

Please cite as:
Eikey EV, Reddy MC, Booth KM, Kvasny L, Blair JL, Li V, Poole ES
Desire to Be Underweight: Exploratory Study on a Weight Loss App Community and User Perceptions of the Impact on Disordered Eating Behaviors
JMIR Mhealth Uhealth 2017;5(10):e150
URL: http://mhealth.jmir.org/2017/10/e150/
doi:10.2196/mhealth.6683
PMID:29025694

Original Paper

Designing a Self-Management App for Young People With Type 1 Diabetes: Methodological Challenges, Experiences, and Recommendations

Pernille Castensøe-Seidenfaden[1], MD; Gitte Reventlov Husted[1], RN, MScN, PhD; Grete Teilmann[1], MD, PhD; Eva Hommel[2], MD, DMSc; Birthe Susanne Olsen[3], MD; Finn Kensing[4], DSc, Professor

[1]Pediatric and Adolescent Department, Nordsjællands Hospital, Hillerød, University of Copenhagen, Hillerød, Denmark

[2]Steno Diabetes Center, Gentofte, Denmark

[3]Pediatric and Adolescent Department, Herlev Hospital, University of Copenhagen, Herlev, Denmark

[4]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Pernille Castensøe-Seidenfaden, MD
Pediatric and Adolescent Department
Nordsjællands Hospital, Hillerød
University of Copenhagen
Dyrehavevej 29, 1521
Hillerød, 3400
Denmark
Phone: 45 29824322
Fax: 45 48293034
Email: pernille.castensoee-seidenfaden@regionh.dk

Abstract

Background: Young people with type 1 diabetes often struggle to self-manage their disease. Mobile health (mHealth) apps show promise in supporting self-management of chronic conditions such as type 1 diabetes. Many health care providers become involved in app development. Unfortunately, limited information is available to guide their selection of appropriate methods, techniques, and tools for a participatory design (PD) project in health care.

Objective: The aim of our study was to develop an mHealth app to support young people in self-managing type 1 diabetes. This paper presents our methodological recommendations based on experiences and reflections from a 2-year research study.

Methods: A mixed methods design was used to identify user needs before designing the app and testing it in a randomized controlled trial. App design was based on qualitative, explorative, interventional, and experimental activities within an overall iterative PD approach. Several techniques and tools were used, including workshops, a mail panel, think-aloud tests, and a feasibility study.

Results: The final mHealth solution was "Young with Diabetes" (YWD). The iterative PD approach supported researchers and designers in understanding the needs of end users (ie, young people, parents, and health care providers) and their assessment of YWD, as well as how to improve app usability and feasibility. It is critical to include all end user groups during all phases of a PD project and to establish a multidisciplinary team to provide the wide range of expertise required to build a usable and useful mHealth app.

Conclusions: Future research is needed to develop and evaluate more efficient PD techniques. Health care providers need guidance on what tools and techniques to choose for which subgroups of users and guidance on how to introduce an app to colleagues to successfully implement an mHealth app in health care organizations. These steps are important for anyone who wants to design an mHealth app for any illness.

(JMIR Mhealth Uhealth 2017;5(10):e124) doi:10.2196/mhealth.8137

KEYWORDS

adolescents; mHealth; diabetes; chronic condition; self-management; transition; participatory design; usability; feasibility; methodological recommendations

Introduction

Type 1 diabetes mellitus (T1DM) is a major health challenge, particularly among young people, who struggle to manage their condition during the transition from childhood to adulthood. Physical, cognitive, and social changes influence their daily T1DM routines (eg, blood glucose measurement, carbohydrate counting, and insulin adjustment). This frequently results in impaired glycemic control [1], increased risk of acute complications [2], and early onset of long-term complications [3,4].

Parents are important supports for young people to self-manage T1DM [5]. However, parents often report frustrations, stress, and worry regarding their role [6]. Schilling et al [7] define self-management for young people with T1DM as a flexible daily process in which young people and their parents share decision making and responsibility for controlling T1DM. The process whereby young people go from being totally dependent on parents to managing their T1DM by themselves is constantly evolving.

Supporting young people in self-managing T1DM is an integral goal of health care [8]. Unfortunately, self-management support can be complex. Health care providers should both guide insulin management and seek insight into young people's lived experiences, such as social life, work, and school, to identify challenges affecting self-management. Furthermore, they must pay attention to young people's needs to develop self-management skills while encouraging and involving parents in supporting their young people [9,10]. Routine care from health care providers appears to have limited effects on self-management and glycemic control [1]. Consequently, supporting young people and parents during the transition from childhood to adulthood is an ongoing challenge for health care providers.

Mobile health (mHealth) apps are promising tools for supporting self-management of a chronic condition such as T1DM [11,12]. They are easily accessible, widely used, and accepted, particularly by young people [13]. They have the potential to improve patient education and enhance communication with health care providers and peers in a convenient and interactive way [14]. Recently, the number of apps to support self-management of chronic conditions such as T1DM in adults has exploded [15]. However, a recent review of mHealth apps for management of chronic physical conditions in adolescents [16] identified only two apps to support adolescents with T1DM [17,18]. Cafazzo et al [17] developed an mHealth app based on interviews with adolescents and their parents to facilitate feedback on blood glucose data. A pilot test (N=20) found an improvement in the frequency of blood glucose monitoring [17]. Frøisland et al [18] tested an mHealth app with a picture-based diabetes diary in addition to a text messaging service in a 3-month pilot study (N=12) and found increased understanding of applied knowledge [18]. Unfortunately, few studies are available, and they are all limited by small sample sizes and the absence of a control group [16]. In addition, mHealth apps are seldom developed on the basis of empirical evidence [19]. In-depth understanding of user needs is often lacking, and the effect of self-management apps is mixed [16,20]. Furthermore, few objective comparisons of methods, tools, and techniques are available to guide health care providers [21] in selecting an appropriate approach for a participatory design (PD) project. Hence, key questions remain unanswered, including which tools and techniques to use and when and where to use them.

This paper presents our methodological recommendations based on experiences and reflections from our 2-year research study. The aim of our study was to develop an mHealth app to support young people in self-managing T1DM.

Methods

The study used a mixed methods design (Figure 1) comprising (1) quantitative and qualitative pre-studies to identify user needs, (2) quantitative and qualitative studies designing the app, and (3) testing of the effect of app use in a randomized controlled trial (RCT).

The focus of this paper is on the process of designing the app, which was based on qualitative, explorative, interventional, and experimental activities within an overall iterative and PD approach [22] (Figure 1). PD promotes user participation in technology design. It enables designers and end users to learn from each other through understanding each other's perspectives and priorities [23]. Involving users in developing an intervention is known to result in higher levels of user acceptance and satisfaction [24] and has previously contributed to changes in the design of mHealth interventions [17,18,25]. Several techniques and tools were used to design the app, including workshops, a mail panel, think-aloud tests, and a feasibility study (Figure 1). Throughout the activities, a purposive sampling strategy [26] was used to ensure variation in age, gender, age at onset of T1DM, and location of diabetes care. Table 1 provides a detailed description of the methods, and Table 2 presents participant characteristics.

A steering group was established to ensure a scalable and usable mHealth app. The group consisted of four physician diabetes team leaders, a professor in computer science, a physician and a nurse with expertise in adolescent medicine, the leader of the telemedicine center in the Capital Region of Denmark, and a consultant from the Danish Agency for Digitisation. The steering group met after the pre-studies and after the app was designed to determine if the study phase was complete [22].

The study was approved by the Danish Data Protection Agency (No. 01980 HIH-2012-013, No. 02249 HIH-2013-016, and No. 04015 NOH-2015-031) and performed in accordance with the ethical recommendations of the Helsinki Declaration. Written consent was obtained from participants and from parents if the young people were under the age of 18 years. Confidentiality and anonymity were assured. Ethical approval of retrospective and qualitative studies by Research Ethics Committee is not necessary in Denmark (No. 15000468, Ref. No. H-15013254).

Figure 1. Mixed methods design.

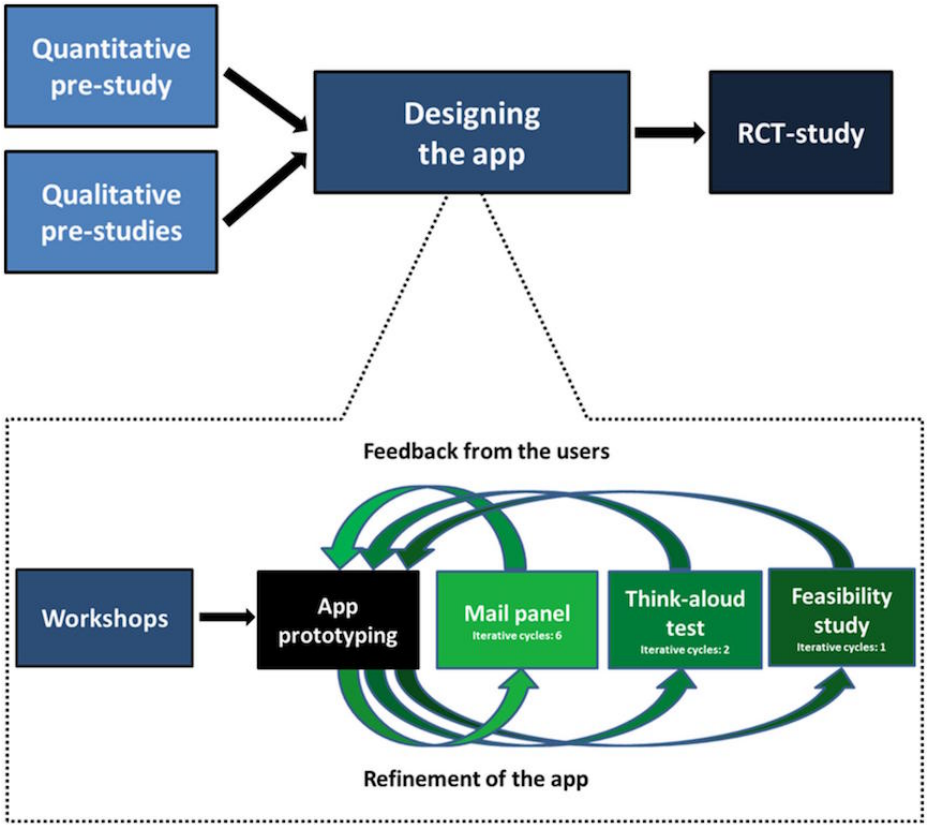


Table 1. Detailed description of app design.

Activity	Aim	Participants	Data collection	Data analysis
Workshops	To develop an app to support young people to self-manage T1DM	Inclusion criteria for young people (and parents): 14-22 years old, T1DM ≥1 year, no psychiatric disorders, in pediatric care or adult care	Workshop themes: My diabetes; App functions; Sensitive topics; Future; To my parents; Knowledge and skills; Design and language	Findings summarized for the IT company to describe functionality and modes of interaction.
		Inclusion criteria for health care providers: on the diabetes team at an adult or a pediatric and adolescent diabetes clinic and with ≥1 year experience working with young people with diabetes	Brainstorming, prioritizing, feedback, and prototyping were recorded and artifacts collected.	
App prototyping	To design the first version of the app	IT company; young people with T1DM and parents; health care providers; experts (interactive design students, illustrator, journalist, movie creator)		
Mail panel	To ensure a reliable and scalable app	See workshop inclusion criteria.	The participants provided written feedback on prototype versions.	Feedback categorized by themes.
Think-aloud test	To ensure a reliable and scalable app	See workshop inclusion criteria. Could not have participated in previous study activities.	Participants "thought aloud" while performing tasks covering the main functions; recordings and observations.	Feedback categorized by themes.
Feasibility study	To test and evaluate the app in a real-life setting	See workshop inclusion criteria; had participated in think-aloud test.	Young people and health care providers tested the app for 5 weeks and completed questionnaires.	Feedback categorized by themes.

Table 2. Participant characteristics.

	Workshops (n=41)			Mail panel (n=49)			Think-aloud test (n=16)			Feasibility study (n=44)	
	YP[a] (n=17)	P[b] (n=10)	HCP[c] (n=14)	YP (n=20)	P (n=3)	HCP (n=26)	YP (n=6)	P (n=4)	HCP (n=6)	YP (n=6)	HCP (n=38)
Age in years, median (SD)	19 (1.7)			19 (2.4)			18 (3.0)			18 (3.0)	
Female, n (%)	11 (65)	6 (60)	13 (93)	12(60)	1 (33)	19 (73)	3 (50)	3 (75)	6 (100)	3 (50)	29 (76)
Pediatric department, n (%)											
Nordsjælland	3 (18)	2 (20)	3 (21)	5 (25)	1 (33)	3 (12)	1 (17)	1 (25)	1 (17)	1 (17)	5 (13)
Herlev	2 (12)	3 (30)	1 (7)	2 (10)	1 (33)	4 (15)	1 (17)	1 (25)	1 (17)	1 (17)	8 (21)
Roskilde	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (12)	1 (17)	1 (25)	1 (17)	1 (17)	5 (13)
Adult department, n (%)											
Hillerød	8 (47)	4 (40)	4 (29)	8 (40)	1 (33)	6 (23)	1 (17)	1 (25)	1 (17)	1 (17)	6 (16)
Steno	4 (24)	1 (10)	6 (43)	5 (25)	0 (0)	8 (31)	1 (17)	0 (0)	1 (17)	1 (17)	9 (24)
Køge	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	2 (8)	1 (17)	0 (0)	1 (17)	1 (17)	5 (13)
Profession, n (%)											
Physician	–	–	4 (29)	–	–	13 (50)	–	–	1 (17)	–	15 (40)
Nurse	–	–	6 (43)	–	–	9 (35)	–	–	3 (50)	–	17 (45)
Dietician	–	–	4 (29)	–	–	4 (15)	–	–	2 (33)	–	6 (16)

[a]YP: young people

[b]P: parents

[c]HCP: health care providers

Quantitative and Qualitative Studies Identifying User Needs

Before designing the app, we explored the needs of young people for self-managing T1DM and possible ways to support these needs. We conducted four main activities.

In a retrospective cohort study (n=126) [27], we found that more than 90% of adolescents had a suboptimal level of hemoglobin A1c around transfer from pediatric to adult care. Those who did not attend clinic visits, whose parents were divorced, or who had a learning disability and/or mental health condition had a higher risk of poor metabolic control.

Using visual storytelling with young people with T1DM (N=9) and their parents (N=13) [28], we explored users' experiences of living with T1DM in individual interviews based on their personal photographs. Young people and their parents experienced the same concerns and challenges related to living with T1DM. They seldom shared these concerns and challenges with each other, which led to misunderstandings, frustration, and conflicts. Four major themes occurred consistently among young people and their parents:

1. Striving for safety. Young people and parents tried to create a "safety net" (ie, hotline, juice, preparing friends) not to risk hypoglycemia. Some adolescents chose to have a high level of blood glucose preventing hypoglycemia, which was supported by some parents.
2. Striving for normality. Young people often felt different from their peers carrying the burden of T1DM. Some tried to be normal by ignoring T1DM, and some parents felt sorry for their child supporting these "breaks." Peers with diabetes helped many feel normal.
3. Striving for independence. Young people and parents longed for the young people to be independent in T1DM management. However, young people faced obstacles such as lack of T1DM knowledge, skills, and parental support. Some avoided clinical visits to hide their incompetence in self-management.
4. Worrying about the future. Both parties worried about the future, such as the risk of long-term complications. Parents thought their child did not worry and chose not to talk about it. However, young people felt alone with their worries, not sharing them with anyone.

In individual interviews with 24 health care providers (10 physicians, 10 nurses, 4 dietitians; unpublished data), we explored health care provider's attitudes towards implementing an app in clinical settings. Two major themes were identified: a new way to collaborate and losing control. All health care providers emphasized that an app could help improve their collaboration with the young people by breaking the ice and helping them investigate young people's real challenges and concerns and by providing ongoing support between clinic visits. On the other hand, health care providers feared losing control of the content of consultations. They feared that their authority would be questioned if they lacked competency with the app

and that its use would be too time-consuming. Health care providers preferred electronic messages to be sent to their work email addresses so they would not have to check two devices.

Finally, we identified relevant security regulations to ensure we met the national standards for login procedures and exchanging messages with peers and health care providers. Compliance with security regulations was assured through six regular meetings throughout the design process with consultants from an IT company and from a public health-technology center.

Designing the App

To develop the mHealth app to support young people in self-managing T1DM, we invited young people with T1DM, their parents, and health care providers to participate in workshops. The first version of the app was developed from the workshop findings.

Workshops

Seven workshops were held in November and December 2014. In total, 17 young people aged 16-21 years, 10 parents, and 14 health care providers participated in one or more workshops (Table 2). In addition, 26 individuals with specialized expertise and knowledge, such as dieticians, psychologists, a social worker, interactive design students, a journalist, information technology (IT) consultants, a telemedicine consultant, and other health care providers with an interest in adolescent medicine, participated in workshops to find ways to meet the needs of the young people and their parents. Each workshop included 11-21 participants (Multimedia Appendix 1).

Workshop content was based on the results from our quantitative and qualitative pre-studies [27,28]. The results from the pre-studies were merged by a mixed methods concurrent design [29] to interpret the challenges that young people and parents face living with T1DM. By applying a mixed methods sequential design [29], the results from the pre-studies were used to inform workshops. As an example, visual storytelling elucidated young people's lack of T1DM knowledge, which was used in Workshop 6 "Knowledge and skills." Workshop themes are listed in Table 1. Each workshop lasted 2½ hours and included a 5-minute introduction, individual brainstorming, prioritizing ideas, and sketching prototypes based on the ideas. Workshop participants were grouped by whether they were young people, parents, or health care providers. The workshops were audio- and video recorded. All input (Post-it notes, flip charts, prototypes, digital records) were collected and categorized by theme; the IT company incorporated the final list of themes into a description of functionality and modes of interaction.

App Prototyping

The IT company built a preliminary version of the app on iOS and Android platforms. Interactive design students were invited to improve the app design and create animations. Young people with T1DM created video self-portraits ("selfies") on a variety of topics, such as how to tell peers one has T1DM or their experiences with low blood sugar. A professional illustrator created graphic elements, and a journalist wrote youth-friendly texts and tips. Experienced providers on diabetes teams (physicians, nurses, dieticians, psychologists, and social workers) revised and approved the final app content.

We were unable to fulfill all needs identified by users. For example, we did not add a mentor/mentee function due to the complexity of evaluating this intervention. However, a supplementary Web-based secure messaging function for health care providers was developed to enable contact with young people. Multimedia Appendix 1 provides an overview of the needs addressed in each of the workshops, the main ideas identified, and the resulting functionality and modes of interaction in the app.

We aimed to evaluate and refine the app. Our primary concern was its usability and feasibility; the IT company took responsibility for technical testing. Our evaluation relied on three techniques (mail panel, think-aloud testing, and feasibility study). A mixed methods sequential design [29] connected findings and ideas from one qualitative study in the design phase to the next to iteratively inform the prototyping and refinement of the app.

Mail Panel

A mail panel comprising 20 young people aged 17-26 years, 3 parents, and 26 health care providers gave feedback on screen shots of the app (mock-ups) and first versions of the app (test flights) in six iterative cycles from March to October 2015. The panel received mail with questions and attached mock-ups (Cycles 1-4) or test flights (Cycle 5-6). The questions focused on layout, content, and functions. The feedback was collected, and the IT company refined the app before the next cycle (Multimedia Appendix 2).

Think-Aloud Testing

Think-aloud tests [30] were performed to understand how users experienced the app interface [31]. Six young people aged 15-22 years, 4 parents, and 6 health care providers participated. None of the young people or parents had participated in previous activities. In individual sessions that lasted 15-53 minutes, participants were asked to verbalize their thoughts while performing app-related tasks (Multimedia Appendix 3). Tasks were designed in collaboration with the IT company to test the interface. Participants were prompted if they found thinking aloud challenging (eg, "Tell me what you are thinking"). The sessions were digitally recorded and observed, and comments and nonverbal reactions were noted. Data were categorized by themes. Based on the feedback, the IT company refined the app before next cycle. A total of two iterative cycles were completed (Cycles 5 and 6; Multimedia Appendix 2) in September and October 2015.

Feasibility Study

A 5-week feasibility study was conducted from October to December 2015 [32,33]. Author PC-S presented the app to a total of 38 health care providers individually (n=13) or in eight groups (n=25). The presentation included a summary of young people's and parents' needs [28], a 15-minute introduction to the app, and two role-playing scenarios in which health care providers were asked to introduce the app and use it as a dialogue ice breaker while a colleague or PC-S played the role

of a young person with T1DM. Finally, health care providers were asked to familiarize themselves with the app in their outpatient clinics and use it with at least 2 young people with T1DM during the 5-week study period. Feasibility was evaluated by an electronic questionnaire sent to health care providers that contained questions such as:

- Does the app work on your device?
- To whom have you introduced the app?
- What challenges have you experienced using the app?
- What has to be changed before conducting an RCT evaluating the app?

Six young people aged 15-22 years who had participated in think-aloud tests were individually introduced to the app in 15-minute sessions and asked to test it for 5 weeks at home and in cooperation with their health care providers and parents. They received a paper questionnaire with questions such as:

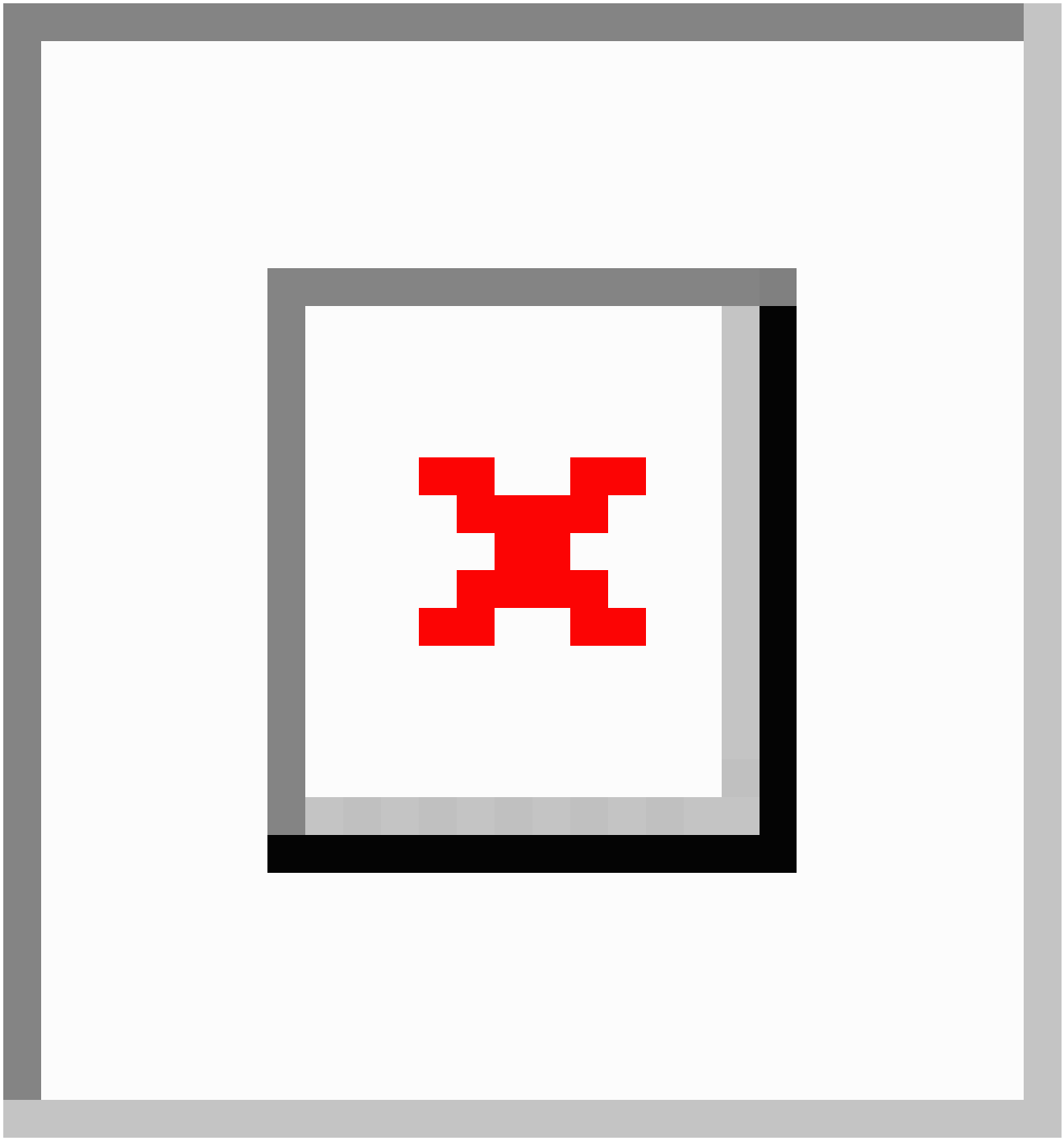
- Has the app helped you?
- How has the app helped you?
- Have you used the app in collaboration with health care provider or parent?
- Would you recommend the app to peers?

The questions were based on previous experiences of the IT company with app development; 2 health care providers and 2 young people assessed face validity before the questionnaire was distributed to test participants. Their responses were categorized into themes, and the IT company made final refinements of the app based on the feedback.

Young With Diabetes App

The final mHealth solution consists of the app "Young with Diabetes" (YWD) and an additional Web-based mail module through which health care providers receive messages from young people. The eight main functions in the app are outlined in Figure 2.

Figure 2. Young with Diabetes.



Results

Our experience yielded valuable learnings and a set of recommendations for future app development (Table 3).

Mixed Methods Design

A mixed methods design allowed us to obtain a nuanced understanding of users' needs and challenges before the workshops [29]. This knowledge was essential to guiding both workshop content and design of the app. In addition, we used a variety of qualitative techniques to explore how users viewed and assessed YWD. This gave us a better understanding of the app content, user interface, and technical issues and was invaluable to further refining the app before implementation. By applying a mixed methods design [29], the findings from the pre-studies informed the design process resulting in a final version of YWD that met the users' needs and challenges. As an example, one of the findings from visual storytelling was "striving for normality." Young people strove for normality in order to not feel different from their peers without T1DM. In addition, meeting peers with T1DM often helped them feel normal. This finding was approached in Workshop 3 "Sensitive topics," where ideas were generated on how to support young people to feel normal. The final app consequently ended up with a chat room and video self-portraits to share experiences with peers, in addition to an information topic on how to tell others that you have T1DM. Furthermore, the parent section

informed parents on how it can actually be to have T1DM including the young peoples' struggle for normality.

Applying a mixed methods design is consistent with a recent design study [34] reporting that qualitative and quantitative results contributed to a comprehensive understanding of the technology and area of concern. On the contrary, a recent PD project [35] developed a patient-centered mHealth app for young people with T1DM and their parents using only qualitative methods, which provided limited information for improving the app. A mixed methods design is highly recommended for future PD projects to gain a comprehensive and nuanced knowledge of the area of concern and meet the needs of users [29].

Participatory Approach

Our PD approach engaged all types of end users (young people, parents, and health care providers) in designing the app. End users were essential collaborators, helping researchers and designers further understand their challenges and needs, generating ideas, giving feedback, and testing the new technology to ensure a usable and feasible product. Similar to previous PD projects [17,18,25], user input contributed to crucial changes in the technology. We found that different types of users had different approaches when engaged in workshop activities. For example, young people generally found it easier to create paper prototypes than did parents, and health care providers and young people often shared new apps with each other, fostering new ideas. We also found it useful to separate workshop participants by user type because they seemed to share ideas more readily than they did in mixed groups. More research is needed to improve our understanding of how best to engage subgroups of participants and which tools and techniques best suit each type of end user.

Giving a voice to end users and designers and respecting their different views is key to success when developing new technology [17,18,24,25]. However, we found it challenging to resolve conflicting views on functionality while taking resources, such as finances and time, into account [21,36]. The seven workshops yielded large amounts of user input and feedback: 46 hours of digital records in addition to several Post-it notes, flip charts, pictures, etc. PC-S, the IT company, and the steering group made final decisions about app content. Taking resources into consideration, they had to eliminate some user-requested functions, such as a mentor/mentee function, a monthly newsfeed, customizing the app, and the ongoing possibility of uploading new video self-portraits. Future app development should consider incorporating these functions, since customization is especially known to be critical for app engagement [35]. In addition, guidelines are needed as to how to manage (collect, analyze, and prioritize) data in future PD projects [36,37]. For example, we were challenged by questions such as: Should digital records be transcribed? How should data be analyzed in PD activities? How do we prioritize and eliminate ideas that are highly valued? Such guidance may help to limit the large amount of data that often challenge PD projects [36] and reduce time-consuming data management activities.

Diverse Team of Experts

Researchers, designers, or end users alone could not have created YWD. The expertise of a diverse team was crucial to building a new health care technology platform and creating the content (eg, information, quizzes, pictures, illustrations, movies, video self-portraits). The authors collaborated with educational institutions such as the Danish School of Media and Journalism to obtain needed expertise at minimal cost and provide students with "real life" projects. In addition, a public-private partnership was established between the hospital and the IT company. The partnership created a team of highly engaged stakeholders who contributed to the study and shared an interest in the study's success. We strongly recommend establishing a multidisciplinary team before developing an mHealth app in health care, which is consistent with previous studies [38]. In addition, it may be beneficial for health care providers, who are often inexperienced in designing new technology, to cooperate with an expert in conducting PD projects to guide the design process. Stakeholders who are considering developing an mHealth app may benefit by meeting with experts, such as innovation and security consultants and lawyers (to draft a partnership agreement) very early in the process. For instance, a group has been established in the capital region of Denmark to guide health care providers through facets of new app studies such as addressing security regulations, facilitating workshops, etc [39].

Table 3. Recommendations for future app development.

Challenges	Experiences	Recommendations	Suggestions for further research
Engaging user subgroups	Participants who are separated in user groups seemed freer to share ideas than they may have been in mixed groups.	Separate participants in user groups in workshops sessions.	What tools and techniques are best suited for different types of end users (ie, young people, parents, health care providers)?
	Young people found it easier to create paper prototypes, compared to parents and health care providers.	Creating paper prototypes is an effective tool for engagement, especially with young people in activities to generate ideas.	
Resolving conflicting views on functionality	The IT company and the steering group made the final decisions about the content of the app.		Incorporate eliminated functions in a future version since users wanted them.
	Functions (mentor/mentee-function, customization, and monthly newsfeed) were eliminated due to lack of resources.		How to prioritize and eliminate user ideas?
Meeting requirements for building an mHealth app	A diverse team of experts was crucial to meeting the challenge of building a new technology platform within health care.	Invite end users, designers, and a diverse team of experts (eg, illustrators, journalists) to participate in workshops.	
		Collaborate with other educational institutions to meet the need for expertise with minimal cost.	
		Establish public-private partnerships to combine resources and ensure engagement from all stakeholders.	
		Consider engaging an innovation consultant to guide the PD process.	
		Set aside enough time to build the app – it always takes more time than expected.	
		Invite users to participate in meetings with the IT company during app building.	
Designing and refining technology in a rapid, low-cost way	Main activities resulted in large amounts of data (eg, 46 hours of digital records from workshops).	Prolong the prototype stage before developing complex expensive technology.	Guidelines are needed on how to collect, analyze, and prioritize data.
	Ongoing user input from iterative cycles helped designers understand user needs and refine the app.	Consider workshops as an ongoing iterative activity in which users give feedback and propose new ideas to prototypes.	More efficient methods, tools, and techniques are needed to meet the rapid development within technology to avoid outdated app versions.
	Expensive technology challenged our ability to meet the users input.	Use living labs to simulate hospital or home settings to try out paper prototypes and explore future ways to use the new technology.	How do we reduce resource (money and time) use?
Improving the user interface	The mail panel functioned as a consulting panel and provided feedback in a short time that improved the app content. The think-aloud tests explored how users assessed the app (ie, navigation, technical errors).	We highly recommend both a mail panel and think-aloud tests in future PD studies, given the valuable input and the low cost and speed of conducting these techniques.	Solicit larger panels using social media (eg, Facebook, Twitter) to comment and share ideas.
	Combining mail panel and think-aloud tests resulted in a substantial reduction of user problems.	Add digital videos and screen records in think-aloud tests to register physical actions, supporting the interpretation of the results.	Introduce a panel in an earlier phase to supplement or replace face-to-face workshops.
Implementing technology in health care	Interviews with health care providers helped us understand barriers to introducing new technology.	Include end users in all phases of a PD project to ensure adoption.	How to teach health care providers to use new technology in collaboration with young people and parents?

Challenges	Experiences	Recommendations	Suggestions for further research
	Workshops, mail panel, think-aloud tests, and feasibility study helped us to ensure a user-friendly app.	Feasibility test new technology prior to implementation.	
	The feasibility study revealed implementation barriers.	Provide a hotline in case of technical difficulties.	
		Teach health care providers how to use the technology prior to test.	

Iterations

Iterative cycles were introduced based on the feedback from the mail panel, think-aloud tests, and the feasibility study. However, the expensive nature of the technology, which had been developed after workshops before the iterative cycles began, challenged our ability to address user input that arose during iterative cycles after app development. Future studies should prolong the prototype stage. Workshops could be considered as an ongoing iterative activity to enable users to give feedback on mock-ups and propose new ideas. In addition, living labs [40] and scenarios could be used to simulate a hospital or home setting where users could explore paper prototypes early in the design process, reflecting on ways to use the new technology [23]. Living labs have previously proved useful to making quick adjustments to prototypes [41] before developing complex, expensive technology [23]. Currently, a need exists to evaluate new approaches and explore more efficient methods, tools, and techniques for rapid technology development.

Mail Panel and Think-Aloud Tests

YWD was evaluated by a mail panel and think-aloud tests before the feasibility test. The mail panel provided valuable user feedback in a short time frame, such as when researcher and designers were uncertain about the front-page design of YWD (Multimedia Appendix 2). The panel primarily improved the app content by reporting incorrect or missing information, misspellings, and information overload, whereas the think-aloud tests largely explored how users assessed the app interface (navigation, technical errors, and layout). As an example, think-aloud test participants were not able to locate the "tips package" located in the reminder function. This resulted in the creation of a separate tips package function. Similarly, think-aloud test participants perceived the icon illustrating the carbohydrate-counting quiz score as a download symbol and began to wait; the icon was subsequently changed. Combining the mail panel and think-aloud tests made it possible to improve the app's usability [36]. In keeping with previous usability studies, we found a substantial reduction in user-identified problems between iterations [21]. Given the valuable input and the low cost and speed with which these techniques can be used, we highly recommend both mail panel and think-aloud tests in future PD projects. In think-aloud tests, we recommend adding digital videos and screen recordings to register physical actions, making the analyses more objective [21] and supporting designers' interpretation of the results. Technology evaluations may be expanded by engaging a larger panel via social media such as Facebook and Twitter, thus disseminating ideas in a viral way and accomplishing iterations quickly [42]. Finally, it could be interesting to introduce a panel in an earlier study phase to supplement or replace face-to-face workshops in the process of generating new ideas. PD practitioners are exploring these possibilities to facilitate participation and further adoption of new technology [23].

Feasibility Testing

Finally, the 5-week feasibility study evaluated YWD use in real-life settings. Young people's and health care providers' attitudes towards the app were explored, revealing practical and technological challenges. These challenges would otherwise have been apparent only after implementation. Young people found the app informative and found that it provided them with a range of self-management support, such as the opportunity to write to their health care providers. They all reported that they would recommend the app to peers. In addition, health care providers described the app as both intuitive to use and relevant to collaborating with young people with T1DM. However, some technical difficulties were reported regarding screen setup and unstable wireless access networks at diabetes clinics, and sending or receiving messages did not always work. None of the young people initiated messages to peers or created notes. To get more activity in the "Chat Room," young people suggested more participants, notifications about new messages, and input from a moderator. Testing YWD's feasibility in a "real" clinical setting aligns with Medical Research Council guidance for the evaluation of complex interventions [43]. We found the feasibility study to be invaluable to further adaptation of YWD; we improved wireless access at clinics, addressed technical issues, and introduced message notifications. We hope that feasibility testing will help prevent challenges and frustrations that often follow the introduction of new health care technology.

Implementation in Real-Life Settings

To facilitate successful implementation, we explored health care providers' perspectives during interviews before designing the app; addressing their concerns was "the first step in embedding the system into practice" (p. 573 [25]). However, during the feasibility study, only half of the providers (n=19) introduced the app to at least one person (young people, colleague, or family). Of these, 14 providers successfully introduced the app to one or more young people. Reasons that health care providers did not introduce the app included a lack of eligible patients, limited time, unstable wireless access, or forgetting to do so. To ensure that providers were able to use the app, we asked them to rate themselves on a readiness scale from 1-10, with 1 representing not ready and 10 representing

absolutely ready [44]. Nearly half (16/38, 42%) of providers rated themselves lower than 7 (median 7, range 1-10). This is consistent with previous studies documenting challenges in adoption of new health care technology [45]. Several explanations should be considered. First, the introduction of new technology may cause a disruptive change in providers' usual workflow [46] and the benefits of the technology may take time to materialize [47]. Our interviews with health care providers (unpublished) revealed that some feared that their authority would be questioned if they were not fully competent at using the app. The training session included "hands-on" activities to simulate real-time scenarios, as recommended [48,49]. However, we may not have spent enough time practicing the scenarios or could have used another teaching approach. In addition, lack of competency at app use may have influenced health care providers' use of YWD, as identified in other studies [50,51], making it challenging for them to engage effectively with young people via the app. Furthermore, some providers were challenged by technical issues related to wireless access and the app itself. Finally, the extensive supporting material (23 informational articles, 43 video selfies, 3 videos of the adult department, 4 animations, several tips packages) required health care providers to spend many hours becoming familiar with YWD. Time should be allocated for health care providers to become familiar with a new technology before introducing it into their practice.

Future implementation of YWD will require regular updates and ongoing staff training to address challenges. For instance, some providers may need more training due to lack of eHealth skills [52,53]. In addition, there is a need for continuous app support via a hotline to overcome technical challenges and wireless access barriers [54,55] and to provide a smooth transition to a new workflow and success in actual usage [50,56]. Health care providers often function as gatekeepers deciding which patients they believe the technology will work for [57]. It could be interesting to consider mHealth apps as "prescriptions" in the future, since prescriptions are often seen as more "serious" recommendations, hopefully enhancing the use of technology. Our feasibility study was followed by health care providers spending more time studying the app, additional training in introducing the app to young people, an app refresher session, help reinstalling the app, frequent app support visits from PC-S, and establishment of a hotline for technical support. Following these additional interventions, all providers rated themselves at 7 or higher on the readiness scale.

Discussion

Strengths and Limitations

The strengths of our study are the mixed methods design and rigorous PD approach applied to ensure that the app would be relevant to all groups of end users and the evaluation of YWD in real-life settings [58]. However, time and financial resources limited our ability to fulfill all user needs. We invited users to participate in more than one activity, which may have biased our results toward favoring their preferences and perhaps made it more difficult for them to continue to think critically about the study. To mitigate this risk, we invited a new set of end users to participate in think-aloud tests and feasibility studies. However, one may argue that participation in more than one main activity is required to gain a mutual understanding of the design process. Due to limited time and economic resources, no parents were included in the feasibility study, thus we do not know if the app is feasible or suitable for them. This shortcoming should be approached by including parents in future tests of YWD. We were unaware of the relevance of including users in the initial meetings with the IT company during the design of the first versions of YWD. We did not use validated questionnaires to score a wider variety of concepts related to eHealth literacy [59], such as functionality, modes of interaction, and user acceptance [60] since these questionnaires were not available in Danish at the time. Doing so would be preferable in future studies. We tested the app over a relatively short period of time and thus do not know its long-term impact or the likelihood that users will stop using it over time [61,62]. Finally, PC-S performed all the main activities. This may have influenced participants' reflections during think-aloud tests and their evaluation of the feasibility study, affecting the resulting mHealth solution in unknown ways.

Conclusion

Our study is an important step for any stakeholder who wants to design an mHealth app for any illness. It paves the way for future PD projects within health care by underscoring the importance of including end user groups during all phases and establishing a multidisciplinary team to provide the wide range of expertise required to build an mHealth app. Before building expensive app versions, we suggest prolonging the prototype stage through workshops, mail panel, think-aloud tests, or scenarios in living labs. Our study highlights a crucial need to develop and validate more efficient PD tools and techniques and to focus on the tools and techniques that are best for specific user groups. Finally, we need to understand how to successfully and efficiently introduce an app to health care providers before we can succeed in implementing mHealth apps in real-life health care settings. An RCT is currently examining the efficacy of YWD in improving the self-management skills of young people with T1DM by measuring hemoglobin A1c and three psychometric scales: Perceived Competence in Diabetes Scale [63], Health Care Climate Questionnaire [63], and Problem Areas in Diabetes care survey [64]. The RCT is followed up by individual interviews qualitatively evaluating YWD. If it proves effective, YWD may potentially serve as a model for supporting self-management and transitions for young people with other chronic or long-term conditions.

Acknowledgments

A special thanks to the young people, parents, and health care providers who participated in this study for kindly sharing their ideas and thoughts with us and spending time testing the technology. Also, thanks to diabetes psychologist Clea Bruun Jacobsen,

dietitian Camilla Storm-Frydenberg, illustrator Mette Ehlers, and interactive design students attending their second year at Danish School of Media and Journalism. Thanks to the Unit of Research and Innovation, Capital Region of Denmark, for kindly sharing ideas about facilitating workshops, and Centre for IT, Medical Technology and Telephony Services, Capital Region of Denmark, and Center of Telemedicine for continuous support in regard to regulatory issues. A thank you to journalist Lene Finnemann for writing inspiring texts, to Jesper Holm for creating the movies of the adult departments, and to Copenhagen Movie and Photo School for editing the video self-portraits. Finally, thanks to the IT enterprise, Mobile Fitness, for partnership during the process.

This study was funded by a research grant from Center of Telemedicine, Capital Region, Denmark, a research grant from Nordsjællands Hospital, University of Copenhagen, Rosalie Petersen's Foundation, Olga Bryde Nielsen's Foundation, and Regionrådet's Foundation for Particularly Vulnerable Citizens, Capital Region, Denmark.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Workshop overview.

[PDF File (Adobe PDF File), 238KB - mhealth_v5i10e124_app1.pdf]

Multimedia Appendix 2

Description of the iterative cycles.

[PDF File (Adobe PDF File), 633KB - mhealth_v5i10e124_app2.pdf]

Multimedia Appendix 3

Think-aloud tasks.

[PDF File (Adobe PDF File), 35KB - mhealth_v5i10e124_app3.pdf]

References

1. Petitti DB, Klingensmith GJ, Bell RA, Andrews JS, Dabelea D, Imperatore G, SEARCH for Diabetes in Youth Study Group. Glycemic control in youth with diabetes: the SEARCH for diabetes in Youth Study. J Pediatr 2009 Nov;155(5):668-672, e661-663 [FREE Full text] [doi: 10.1016/j.jpeds.2009.05.025] [Medline: 19643434]
2. Nakhla M, Daneman D, To T, Paradis G, Guttmann A. Transition to adult care for youths with diabetes mellitus: findings from a Universal Health Care System. Pediatrics 2009 Dec;124(6):e1134-e1141. [doi: 10.1542/peds.2009-0041] [Medline: 19933731]
3. Bryden KS, Dunger DB, Mayou RA, Peveler RC, Neil HAW. Poor prognosis of young adults with type 1 diabetes: a longitudinal study. Diabetes Care 2003 Apr;26(4):1052-1057. [Medline: 12663572]
4. Wills CJ, Scott A, Swift PGF, Davies MJ, Mackie ADR, Mansell P. Retrospective review of care and outcomes in young adults with type 1 diabetes. BMJ 2003 Aug 02;327(7409):260-261 [FREE Full text] [doi: 10.1136/bmj.327.7409.260] [Medline: 12896937]
5. Hanna KM, Guthrie D. Parents' and adolescents' perceptions of helpful and nonhelpful support for adolescents' assumption of diabetes management responsibility. Issues Compr Pediatr Nurs 2001;24(4):209-223. [Medline: 11769207]
6. Haugstvedt A, Wentzel-Larsen T, Graue M, Søvik O, Rokne B. Fear of hypoglycaemia in mothers and fathers of children with Type 1 diabetes is associated with poor glycaemic control and parental emotional distress: a population-based study. Diabet Med 2010 Jan;27(1):72-78. [doi: 10.1111/j.1464-5491.2009.02867.x] [Medline: 20121892]
7. Schilling LS, Dixon JK, Knafl KA, Lynn MR, Murphy K, Dumser S, et al. A new self-report measure of self-management of type 1 diabetes for adolescents. Nurs Res 2009;58(4):228-236 [FREE Full text] [doi: 10.1097/NNR.0b013e3181ac142a] [Medline: 19561555]
8. Harvey PW, Petkov JN, Misan G, Fuller J, Battersby MW, Cayetano TN, et al. Self-management support and training for patients with chronic and complex conditions improves health-related behaviour and health outcomes. Aust Health Rev 2008 May;32(2):330-338. [Medline: 18447824]
9. Glasgow RE, Davis CL, Funnell MM, Beck A. Implementing practical interventions to support chronic illness self-management. Jt Comm J Qual Saf 2003 Nov;29(11):563-574. [Medline: 14619349]
10. Lugasi T, Achille M, Stevenson M. Patients' perspective on factors that facilitate transition from child-centered to adult-centered health care: a theory integrated metasummary of quantitative and qualitative studies. J Adolesc Health 2011 May;48(5):429-440. [doi: 10.1016/j.jadohealth.2010.10.016] [Medline: 21501800]

11. Franklin VL, Greene A, Waller A, Greene SA, Pagliari C. Patients' engagement with “Sweet Talk” - a text messaging support system for young people with diabetes. J Med Internet Res 2008 Jun 30;10(2):e20 [FREE Full text] [doi: 10.2196/jmir.962] [Medline: 18653444]
12. 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]
13. Media Literacy Clearinghouse. 2013. Media Use Statistics: Resources on Media Habits of Children, Youth and Adults URL: http://frankwbaker.com/mlc/media-use-statistics/ [accessed 2017-08-11] [WebCite Cache ID 6siXb8kzA]
14. Griffiths F, Bryce C, Cave J, Dritsaki M, Fraser J, Hamilton K, et al. Timely Digital Patient-Clinician Communication in Specialist Clinical Services for Young People: A Mixed-Methods Study (The LYNC Study). J Med Internet Res 2017 Apr 10;19(4):e102 [FREE Full text] [doi: 10.2196/jmir.7154] [Medline: 28396301]
15. Research 2 Guidance. The Market for mHealth App Services Will Reach $26 Billion by 2017 URL: https://research2guidance.com/the-market-for-mhealth-app-services-will-reach-26-billion-by-2017/ [accessed 2017-08-11] [WebCite Cache ID 6sdUqFVQb]
16. 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 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
17. 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 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
18. Frøisland DH, Arsand E, Skårderud F. Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study. J Med Internet Res 2012;14(4):e111 [FREE Full text] [doi: 10.2196/jmir.2155] [Medline: 22868871]
19. Buijink AWG, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
20. Or CKL, Tao D. Does the use of consumer health information technology improve outcomes in the patient self-management of diabetes? A meta-analysis and narrative review of randomized controlled trials. Int J Med Inform 2014 May;83(5):320-329. [doi: 10.1016/j.ijmedinf.2014.01.009] [Medline: 24534118]
21. Kushniruk A, Nøhr C. Participatory Design, User Involvement and Health IT Evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
22. Kensing F, Sigurdardottir H, Stoop A. MUST--a participatory method for designing sustainable health IT. Stud Health Technol Inform 2007;129(Pt 2):1204-1208. [Medline: 17911906]
23. Simonsen J, Robertson T. Routledge International Handbook of Participatory Design. London, UK: Routledge; 2012.
24. Shneiderman B, Plaisant C. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 5th ed. Boston, MA: Addison-Wesley; 2010.
25. Aldiss S, Taylor R, Soanes L, Maguire R, Sage M, Kearney N, et al. Working in collaboration with young people and health professionals. A staged approach to the implementation of a randomised controlled trial. Journal of Research in Nursing 2010 Sep 17;16(6):561-576. [doi: 10.1177/1744987110380803]
26. Patton M. Qualitative Evaluation and Research Methods. 2nd. ed. Thousand Oaks, CA: SAGE Publications; 1990.
27. Castensøe-Seidenfaden P, Jensen AK, Smedegaard H, Hommel E, Husted GR, Pedersen-Bjergaard U, et al. Clinical, behavioural and social indicators for poor glycaemic control around the time of transfer to adult care: a longitudinal study of 126 young people with diabetes. Diabet Med 2017 May;34(5):667-675. [doi: 10.1111/dme.13318] [Medline: 28099760]
28. Castensøe-Seidenfaden P, Teilmann G, Kensing F, Hommel E, Olsen BS, Husted GR. Isolated thoughts and feelings and unsolved concerns: adolescents' and parents' perspectives on living with type 1 diabetes - a qualitative study using visual storytelling. J Clin Nurs 2016 Nov 16. [doi: 10.1111/jocn.13649] [Medline: 27865017]
29. Creswell J, Plano CV. Designing and Conducting Mixed Methods Research. 2nd ed. Thousand Oaks, CA: SAGE Publications; 2011.
30. Nielsen J. The #1 Usability Tool. Thinking Aloud URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [accessed 2017-08-11] [WebCite Cache ID 6sdV2c5lm]
31. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
32. Faulkner MS. Intervention fidelity: ensuring application to practice for youth and families. J Spec Pediatr Nurs 2012 Jan;17(1):33-40 [FREE Full text] [doi: 10.1111/j.1744-6155.2011.00305.x] [Medline: 22188270]
33. Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, Treatment Fidelity Workgroup of the NIH Behavior Change Consortium. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004 Sep;23(5):443-451. [doi: 10.1037/0278-6133.23.5.443] [Medline: 15367063]
34. Yingling LR, Brooks AT, Wallen GR, Peters-Lawrence M, McClurkin M, Cooper-McCann R, et al. Community Engagement to Optimize the Use of Web-Based and Wearable Technology in a Cardiovascular Health and Needs Assessment Study:

A Mixed Methods Approach. JMIR Mhealth Uhealth 2016 Apr 25;4(2):e38 [FREE Full text] [doi: 10.2196/mhealth.4489] [Medline: 27113680]
35. Holtz BE, Murray KM, Hershey DD, Dunneback JK, Cotten SR, Holmstrom AJ, et al. Developing a Patient-Centered mHealth App: A Tool for Adolescents With Type 1 Diabetes and Their Parents. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e53 [FREE Full text] [doi: 10.2196/mhealth.6654] [Medline: 28428167]
36. Kujala S. User involvement: A review of the benefits and challenges. Behaviour & Information Technology 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
37. Wixon D, Ramey J. Field Methods Casebook for Software Design. New York, NY: Wiley; 1996.
38. Clemensen J, Larsen SB, Kyng M, Kirkevold M. Participatory design in health sciences: Using cooperative experimental methods in developing health services and computer technology. Qual Health Res 2007 Jan;17(1):122-130. [doi: 10.1177/1049732306293664] [Medline: 17170250]
39. Region Hovedstaden. 2014. App-gruppe sikrer fælles koordinering of app-indsatsen (App-group ensures joint coordination of app action) URL: https://www.regionh.dk/til-fagfolk/forskning-og-innovation/Nyheder-om-forskning-og-innovation/Sider/2014/App-kompetencegruppe-sikrer-faelles-koordinering-af-app-indsatsen.aspx [accessed 2017-08-11] [WebCite Cache ID 6sdVJnqsZ]
40. Andersen TO, Bansler JP, Kensing F, Moll J. From Prototype to Product: Making Participatory Design of mHealth Commercially Viable. Stud Health Technol Inform 2017;233:95-112. [Medline: 28125417]
41. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2016 Jan 01. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
42. Pilemalm S, Timpka T. Third generation participatory design in health informatics--making user participation applicable to large-scale information system projects. J Biomed Inform 2008 Apr;41(2):327-339 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.004] [Medline: 17981514]
43. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655 [FREE Full text] [Medline: 18824488]
44. Lawrence W, Black C, Tinati T, Cradock S, Begum R, Jarman M, et al. 'Making every contact count': Evaluation of the impact of an intervention to train health and social care practitioners in skills to support health behaviour change. J Health Psychol 2016 Feb;21(2):138-151 [FREE Full text] [doi: 10.1177/1359105314523304] [Medline: 24713156]
45. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health Care Provider Adoption of eHealth: Systematic Literature Review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
46. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
47. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in “early adopter” hospitals. BMJ 2011 Oct 17;343:d6054 [FREE Full text] [Medline: 22006942]
48. Dagroso D, Williams PD, Chesney JD, Lee MM, Theoharis E, Enberg RN. Implementation of an obstetrics EMR module: overcoming user dissatisfaction. J Healthc Inf Manag 2007;21(1):87-94. [Medline: 17299930]
49. Sicotte C, Paré G, Moreault M, Paccioni A. A risk assessment of two interorganizational clinical information systems. J Am Med Inform Assoc 2006;13(5):557-566 [FREE Full text] [doi: 10.1197/jamia.M2012] [Medline: 16799130]
50. Halamka J, Aranow M, Ascenzo C, Bates DW, Berry K, Debor G, et al. E-Prescribing collaboration in Massachusetts: early experiences from regional prescribing projects. J Am Med Inform Assoc 2006;13(3):239-244 [FREE Full text] [doi: 10.1197/jamia.M2028] [Medline: 16501174]
51. Kemper AR, Uren RL, Clark SJ. Adoption of electronic health records in primary care pediatric practices. Pediatrics 2006 Jul;118(1):e20-e24. [doi: 10.1542/peds.2005-3000] [Medline: 16818534]
52. Hing E, Hsiao C. Electronic medical record use by office-based physicians and their practices: United States, 2007. Natl Health Stat Report 2010 Mar 31(23):1-11 [FREE Full text] [Medline: 20632518]
53. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
54. Pagliari C, Donnan P, Morrison J, Ricketts I, Gregor P, Sullivan F. Adoption and perception of electronic clinical communications in Scotland. Inform Prim Care 2005;13(2):97-104 [FREE Full text] [Medline: 15992494]
55. Cherry B, Carter M, Owen D, Lockhart C. Factors affecting electronic health record adoption in long-term care facilities. J Healthc Qual 2008;30(2):37-47. [Medline: 18411891]
56. Fullerton C, Aponte P, Hopkins R, Bragg D, Ballard DJ. Lessons learned from pilot site implementation of an ambulatory electronic health record. Proc (Bayl Univ Med Cent) 2006 Oct;19(4):303-310 [FREE Full text] [Medline: 17106488]
57. Whitten P, Holtz B. Provider utilization of telemedicine: the elephant in the room. Telemed J E Health 2008 Nov;14(9):995-997. [doi: 10.1089/tmj.2008.0126] [Medline: 19035815]
58. Lynch T, Gregor S. User participation in decision support systems development: Influencing system outcomes. Eur J Inf Syst 2004 Nov 02;13(4):286-301. [doi: 10.1057/palgrave.ejis.3000512]

59. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
60. Lyles CR, Sarkar U, Osborn CY. Getting a technology-based diabetes intervention ready for prime time: a review of usability testing studies. Curr Diab Rep 2014 Oct;14(10):534. [doi: 10.1007/s11892-014-0534-9] [Medline: 25173689]
61. Devito DA, Song M, Hawkins R, Aubrecht J, Kovach K, Terhorst L, et al. An intervention fidelity framework for technology-based behavioral interventions. Nurs Res 2011;60(5):340-347 [FREE Full text] [doi: 10.1097/NNR.0b013e31822cc87d] [Medline: 21878796]
62. DeVito DA, Dew MA, Myers B, Begey A, Hawkins R, Ren D, et al. Evaluation of a hand-held, computer-based intervention to promote early self-care behaviors after lung transplant. Clin Transplant 2009;23(4):537-545 [FREE Full text] [doi: 10.1111/j.1399-0012.2009.00992.x] [Medline: 19473201]
63. Williams GC, Freedman ZR, Deci EL. Supporting autonomy to motivate patients with diabetes for glucose control. Diabetes Care 1998 Oct;21(10):1644-1651. [Medline: 9773724]
64. Polonsky WH, Anderson BJ, Lohrer PA, Welch G, Jacobson AM, Aponte JE, et al. Assessment of diabetes-related distress. Diabetes Care 1995 Jun;18(6):754-760. [Medline: 7555499]

Abbreviations

IT: information technology
PD: participatory design
RCT: randomized controlled trial
T1DM: type 1 diabetes
YWD: Young with Diabetes app

Edited by G Eysenbach; submitted 02.06.17; peer-reviewed by B Holtz, K Murray; comments to author 21.06.17; revised version received 30.06.17; accepted 30.06.17; published 23.10.17.

Please cite as:
Castensøe-Seidenfaden P, Reventlov Husted G, Teilmann G, Hommel E, Olsen BS, Kensing F
Designing a Self-Management App for Young People With Type 1 Diabetes: Methodological Challenges, Experiences, and Recommendations
JMIR Mhealth Uhealth 2017;5(10):e124
URL: http://mhealth.jmir.org/2017/10/e124/
doi:10.2196/mhealth.8137
PMID:29061552

Original Paper

Client-Focused Security Assessment of mHealth Apps and Recommended Practices to Prevent or Mitigate Transport Security Issues

Jannis Müthing[1], BSc; Thomas Jäschke[1,2], PhD; Christoph M Friedrich[1], PhD

[1]Department of Computer Science, University of Applied Sciences and Arts Dortmund, Dortmund, Germany

[2]Department of Business Information Systems, FOM University of Applied Sciences, Essen, Germany

Corresponding Author:
Christoph M Friedrich, PhD
Department of Computer Science
University of Applied Sciences and Arts Dortmund
Emil-Figge Str. 42
Dortmund, 44227
Germany
Phone: 49 231755 ext 6796
Email: christoph.friedrich@fh-dortmund.de

Abstract

Background: Mobile health (mHealth) apps show a growing importance for patients and health care professionals. Apps in this category are diverse. Some display important information (ie, drug interactions), whereas others help patients to keep track of their health. However, insufficient transport security can lead to confidentiality issues for patients and medical professionals, as well as safety issues regarding data integrity. mHealth apps should therefore deploy intensified vigilance to protect their data and integrity. This paper analyzes the state of security in mHealth apps.

Objective: The objectives of this study were as follows: (1) identification of relevant transport issues in mHealth apps, (2) development of a platform for test purposes, and (3) recommendation of practices to mitigate them.

Methods: Security characteristics relevant to the transport security of mHealth apps were assessed, presented, and discussed. These characteristics were used in the development of a prototypical platform facilitating streamlined tests of apps. For the tests, six lists of the 10 most downloaded free apps from three countries and two stores were selected. As some apps were part of these top 10 lists in more than one country, 53 unique apps were tested.

Results: Out of the 53 apps tested from three European App Stores for Android and iOS, 21/53 (40%) showed critical results. All 21 apps failed to guarantee the integrity of data displayed. A total of 18 apps leaked private data or were observable in a way that compromised confidentiality between apps and their servers; 17 apps used unprotected connections; and two apps failed to validate certificates correctly. None of the apps tested utilized certificate pinning. Many apps employed analytics or ad providers, undermining user privacy.

Conclusions: The tests show that many mHealth apps do not apply sufficient transport security measures. The most common security issue was the use of any kind of unprotected connection. Some apps used secure connections only for selected tasks, leaving all other traffic vulnerable.

(JMIR Mhealth Uhealth 2017;5(10):e147) doi:10.2196/mhealth.7791

KEYWORDS

mobile health; mobile apps; data security; computer security; confidentiality; health information technology

Introduction

Mobile Health Apps

With the emergence of smartphones, ubiquitous Internet access and the app ecosystems around, health information technology also found its way to these devices. Mobile health (mHealth) describes using mobile devices to facilitate medical or health-related purposes [1]. Among many other apps, mHealth apps may offer a means of communication between patients and medical professionals. They also give patients the ability to keep track of their medical characteristics [2-5].

In developing countries, smartphones are often the only means of Internet access. mHealth apps on smartphones can thus help to minimize discrepancies in health care worldwide [6,7]. Because they are used in a diverse set of medical apps, they have a heightened need for protection [8]. To offer any security, device vendors must ensure fast security patches for smartphones. This represents an issue especially for low-cost Android-based devices [9]. Beyond device security, security of data in transport is relevant and will be the focus of this paper.

European privacy regulations set an additional baseline for data handling by app providers [10]. The regulations are binding in European countries only. The Privacy Code of Conduct on mHealth apps by the European Commission represents an important initiative outlining the heightened security requirements for mHealth apps [11].

Studies have shown that there is an existing concern about information security [12,13]. mHealth-related apps that do not provide appropriate security might impede the growth of the sector.

Transport Security

To provide information or to enable the transmission of (medical) data to a service provider, an app must communicate with servers. As soon as data are sent through public infrastructure, data can potentially be observed, modified, or redirected. Without any protection, this endangers the integrity of data displayed, gives away potentially sensitive data, and enables malicious parties to impersonate the victim.

The transport layer security (TLS) protocol makes up the foundation of the modern Internet's security infrastructure. It was designed to give protection against the aforementioned problems, offering authentication, data integrity, and confidentiality through asymmetric and symmetric cryptography. In the recent past, protocol weaknesses such as Padding Oracle On Downgraded Legacy Encryption [14], Browser Exploit Against SSL and TLS, Factoring RSA Export Keys, and others, as well as implementation problems such as Heartbleed [15,16] and Apple's goto fail bug [17] have arisen. The use of older protocol versions or deprecated implementations can lead to these or other issues surfacing and compromising the security and privacy of users.

Some prior research examined app source code for transport security issues using static code analysis [18-22], showing clearly that many apps are not using aforementioned up-to-date security measurements and consequently putting users at risk. The methods used in this paper will rely on the observation of communication between the client app and servers, and thus enabling observations under real-world conditions. Consequently, the research presented also does not focus on the analysis of data locally stored on a smartphone [23]. Other transport security issues relevant to this research are listed as part of the Open Web Application Security Project Mobile Top 10 [24].

Apps on mobile devices conceal details of communication with their servers from end users. Whereas a user of a website might be able to identify a website as insecure and be warned about certificate issues, a mobile app does not automatically warn the user about invalid certificates or missing encryption [25]. This highlights the importance of independent evaluation of mobile app transport security.

Prior Work

In existing research, metadata of mHealth apps on iOS and Android app stores were analyzed and evaluated [26]. No test or technical analysis was performed in that publication. Other research focused on health-related apps in Chinese App Stores [27]. This paper also did a comprehensive metadata analysis and a manual screening of popular Chinese mobile apps [27]. The security analysis is limited to viewing of documentation or auditing report availability from the app's developer. The paper did find that information security was absent in 97% of the evaluated apps [27].

Furthermore, a framework for risk assessment of mHealth apps was proposed [8]. The research focuses on evaluation and categorization criteria for apps and represents an excellent motivation for this work.

In other existing literature, a study on security aspects of Android apps was performed, taking an in-depth look at 22 mHealth apps [28]. Here data in transit as well as device data (on Secure Digital cards or in system log files) were considered to evaluate the apps. Their results contain the finding that 18 of these apps send data unencrypted over the Internet.

Beyond the field of health-related security analysis, Gagnon et al proposed the AndroSSL Platform to test Android apps regarding transport security [29]. The approach presented here was to test apps in an Android virtual device, utilizing a virtual test bed for Android apps [30]. This enabled to record a test once and repeat it multiple times automatically. The focus was on certificate validation when secure connections were used. By being able to repeat a test automatically, it was possible to issue different secure sockets layer (SSL) certificates to find out whether the client validated them correctly. These or similar test scenarios are also found in other relevant research [22,31]. This led to the incorporation of similar tests into the research presented in this paper.

The primary objective of the research presented in this paper was to assess prominent transport security issues in popular mHealth apps and to outline ways for developers of such apps to mitigate these issues.

The following Methods section will first outline the app selection criteria. A description of all the aspects analyzed during the tests will be given in this section. Subsequently, the system used for the tests will be described. In the Results section, the apps selected for testing by the criteria described before will be given, followed by a description of how the previously described system was applied for testing. The last section discusses common security concerns found during the tests, compares prior work with this paper, and recommends practices to mitigate the security issues discussed.

Methods

App Selection

To achieve appropriate diversity in the test pool, mHealth apps from different European countries were chosen. To mitigate any platform-dependent bias, apps for Android as well as for iOS were tested.

Relevant Transport Security Considerations

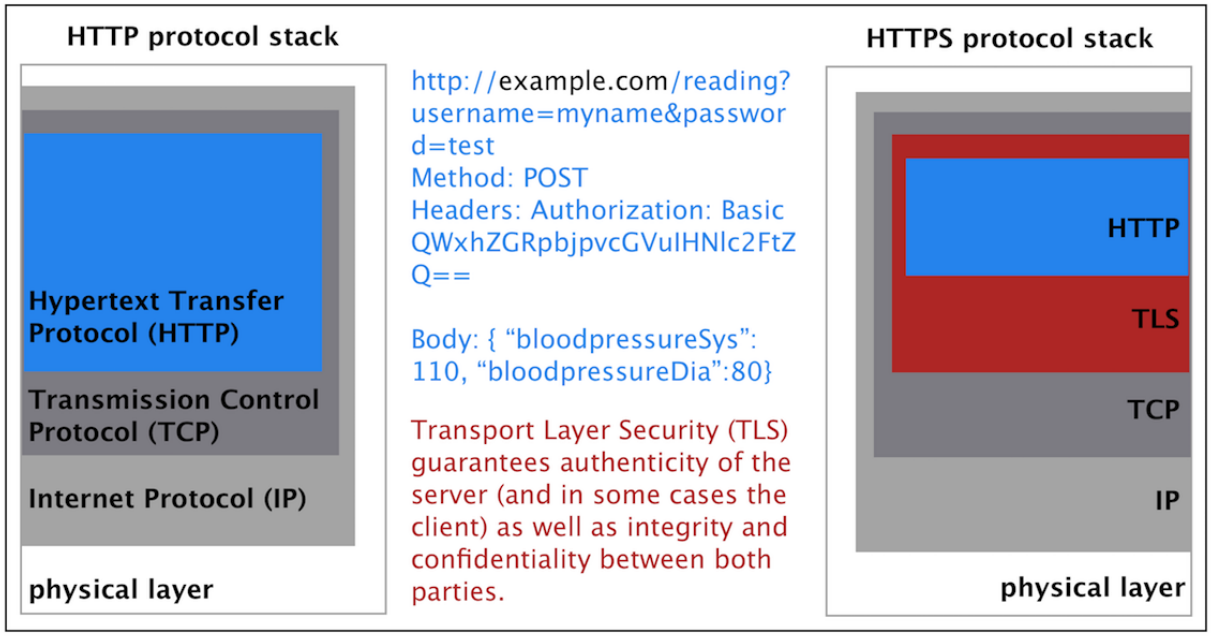
This section will describe each characteristic that will be considered in the tests performed later in this paper.

HTTP (Hypertext Transfer Protocol) is widely used by mobile apps to facilitate server-client communication [32]. This paper focuses on information transmitted utilizing this protocol. HTTP is an application layer protocol (layer 4 in the Transmission Control Protocol and Internet Protocol stack) and can be used on top of a secure TLS connection [33,34]. TLS and its predecessor SSL are designed to ensure confidentiality (encryption), integrity, and authenticity between the parties involved in the communication. The protocol utilizes asymmetric cryptography and a public key infrastructure during its initial handshake and key exchange. Later communication is symmetrically encrypted [35,36]. In Figure 1, the protocol stacks for unprotected HTTP and protected HTTPS (Hypertext Transfer Protocol Secure) are illustrated. The version of the transport security protocol in use is of high relevance to the security of a connection. Earlier versions of TLS and SSL had severe security issues [14,37,38]. This makes testing for the use of HTTPS in general and for the TLS version imperative.

By default, a TLS implementation, for example, in a browser or in a mobile operating system trusts a number of root public certificates from certificate authorities [39]. When an app makes a secure connection to a server, this server authenticates itself with a certificate. The TLS client on the smartphone validates that this certificate was derived from one of its trusted certificates. Because these lists of trusted certificates are not controllable by the app developer, it is possible that it contains compromised certificates. As soon as an app trusts such a rogue certificate, the owner of the rogue certificate can issue valid certificates for any domain visited by the device and can, therefore, pretend to be any server [21]. This enables an attacker to act as a middle man (man-in-the-middle [MitM]) between the client and the server, leading to undermined integrity of server responses and loss of privacy between the client (and thus the user of an app) and the server [40]. In Android version <7, the user can install such a certificate himself. In later versions, a user cannot install additional CA (Certificate Authority) certificates [41]. It should be noted that system integrity is required for the validation of certificate trust chains to work. Android's inconsistent history with system security in the past could make it more likely that an attacker might use unpatched issues to gain privileges on the system and install any certificates wanted or to do further harm [42,43]. A major issue with Android phones is the lack of willingness in phone manufacturers to ship security updates to their adoptions of Android, leading to a high degree of version fragmentation in the market [9,44]. iOS gives the user a way to install a trusted CA certificate manually.

To make sure an app only connects to the correct servers, apps can be shipped including several trusted certificates. When a secure connection is made, the app validates the server certificate against these certificates. As the app bundle is signed by the developer and consecutively by the store operators (Apple and Google, respectively), it cannot be tampered with later [45,46]. This technique is called pinning. Whereas it brings some important advantages, shipped trusted certificates can expire, making the app unable to connect to its servers. It is also possible that the necessity arises for a certificate to be revoked. This will require the app to be repackaged. Pinning is similar to HTTP public key pinning (HPKP) but does not require changes to the server [47].

Figure 1. The Hypertext Transfer Protocol (HTTP) and Hypertext Transfer Protocol Secure (HTTPS) protocol stacks. The topmost layers (transport layer security [TLS] and HTTP itself) are of most interest. The HTTP protocol contains any relevant data sent to or received from the server. Examples for HTTP data are written in blue. These data are readable by any third party when TLS is not used. When HTTP is used on top of TLS, these data are encrypted. Additionally, TLS ensures the integrity of the messages exchanged and the authenticity of the server and in some cases the clients.



Because HPKP depends on a server configuration, it may not prevent all MitM attacks [48]. During the tests, a self-issued CA certificate is utilized and installed on devices used for testing to inspect encrypted traffic in a part of the tests. When a connection attempt is consistently aborted by the client while the proxy is presenting a certificate derived from the aforementioned CA certificate, pinning is likely to be used by the app. Pinning is enabled if no connections to an app's backend can be made through the proxy.

Another set of tests is inspired by Gagnon et al. It consists of several scenarios to test the certificate validation of TLS implementations in apps [29]. Four of these are part of the tests performed in this paper. In every scenario, the proxy serves a different TLS certificate for each domain requested. The certificates are all invalid and should be rejected by the client app under test. These are the characteristics of the certificates served in the scenarios:

1. Correct domain name, signed by an untrusted CA certificate
2. Self-signed certificate for the domain requested
3. Static host name, signed by a trusted CA
4. Self-signed for a static hostname

Because each scenario requires a separate manual test, only these four scenarios were selected [29]. The last scenario in Gagnon et al's paper did not yield any further results and was therefore excluded from the setup.

Next, the leakage of information is considered. Cookies are used by servers to hold session information [49]. They are transmitted as HTTP header fields. If it is possible to reuse an intercepted cookie, the interceptor can hijack a session. Leaked cookies can also reveal user data directly [50]. Cookies should be protected by a secure connection. A secure cookie scheme can also mitigate the issues [51].

Cookies are one way to identify a client to the server. Users can be authenticated by all kinds of tokens or parts of an HTTP request. Therefore, the system to be developed will look for cookie, set-cookie, and authorization headers.

The authorization header field can contain one of multiple possible values of interest. It may leak usernames and passwords [52], OAuth2 Bearer tokens [53], or other sensitive information.

Additionally, the body and URL string of each request and response will be evaluated for any username or password leaks.

Lastly, the server location is relevant, as it has consequences for the jurisdiction applied. As mentioned earlier, servers outside Europe are not under the European privacy regulation.

Development of System for Semiautomatic Tests of Relevant Transport Security Issues

To be able to rapidly and thoroughly test for the issues discussed above, a Web-based app was developed. This Web-based app should enable users to test apps for vulnerabilities while also facilitating more in-depth analysis. The software is called BProxy.

The app was based on the Zed Attack proxy and was started as a fork of version 2.4.3 [54]. The main points of reusing the existing code were the proxy inspection and dynamic certificate–issuing codebase. Changes were made to dynamically modify how certificates for requested domains are issued (to enable the certificate validation scenarios discussed earlier). A representational state transfer (REST) application programming interface was designed to expose automatic creation and control of proxies [55]. Additionally, an HTTP server exposes the Angular2-based user interface. This Web-based app interacts with the REST interface to control the proxy.

The architecture of BProxy was engineered with fast and simple extensibility in mind. Each single transport security consideration was tested by a separated module. Modules can implement interfaces to register for callbacks and influence properties of TLS handshakes (for the TLS certificate validation tests). Multimedia Appendix 1 shows the general software architecture of the tool. Additionally, BProxy has been released as open source software to help reproducibility of the research presented [56].

During each test of an app, the proxy works in sessions. Before the start of each session, the app under test is relaunched. During a session, the user interacts with it. Any registration or log-in actions are repeated.

First, this enables the system to separate domains used by the app from other domains the device might communicate with (background tasks, changing ads displayed in the app). A domain present in more sessions is more likely to be connected to the app under testing. Second, some sessions are used for the certificate validation tests described earlier.

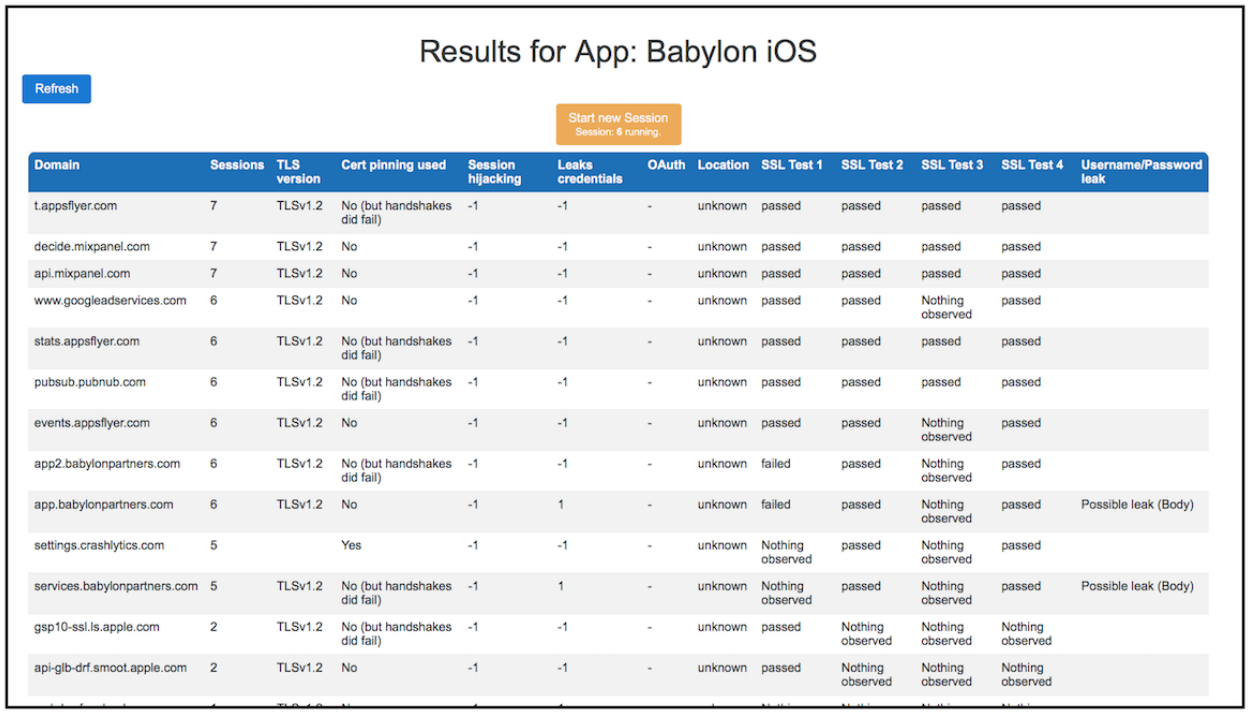
After the necessary number of sessions, a list of domains will be shown. During our tests, two without certificate modifications and four with different certificate validation tests must be run. The results are displayed per domain that the app communicated with. The modules mentioned earlier are responsible for generating these results. Where possible, a user can also display all request and response pairs that are responsible for a certain result displayed. This enables validation of the automatically generated results and further in-depth analysis. The source code for BProxy is available on the Web [56]. An example of how it presents its results is shown in Figure 2.

Limitations

The platform developed as part of the research for this paper should enable even less technology-affine users of mobile apps to conduct tests and get results. These results should give an indication of the value the app's developer assigns to security. As a direct result of the intention of developing such a tool, the choice was made early on to develop it as a Web-based platform. This choice brought certain design limitations. First, the analysis is based on the use of a proxy running on a unique port assigned to the app under test. This proxy can simply be configured on user's devices. It is possible for an app to ignore system proxy setting on Android and iOS, but during all tests, no apps ignored the proxy and any traffic was apparently observable.

As described, the developed system works only semiautomatically. This is to enable tests on apps from the respective app stores on Android as well as on iOS. No research on data locally stored on mobile devices was performed.

Figure 2. BProxy example results output. The columns inform the user about observations made by the proxy: the Transport Layer Security (TLS) version used (TLS version), whether certificate pinning was used (Cert pinning used), whether cookies were observed (Session hijacking), whether authentication tokens were visible (Leaks credentials), if OpenAuthorization (OAuth) tokens were observed (OAuth), the server location for the domain visited (Location), the results for the certificate validation tests (SSL Test 1-4), if usernames or passwords were observed (Username/Password leak). More Information on BProxy's output can be found on the Web.



Domain	Sessions	TLS version	Cert pinning used	Session hijacking	Leaks credentials	OAuth	Location	SSL Test 1	SSL Test 2	SSL Test 3	SSL Test 4	Username/Password leak
t.appsflyer.com	7	TLSv1.2	No (but handshakes did fail)	-1	-1	-	unknown	passed	passed	passed	passed	
decide.mixpanel.com	7	TLSv1.2	No	-1	-1	-	unknown	passed	passed	passed	passed	
api.mixpanel.com	7	TLSv1.2	No	-1	-1	-	unknown	passed	passed	passed	passed	
www.googleadservices.com	6	TLSv1.2	No	-1	-1	-	unknown	passed	passed	Nothing observed	passed	
stats.appsflyer.com	6	TLSv1.2	No (but handshakes did fail)	-1	-1	-	unknown	passed	passed	passed	passed	
pubsub.pubnub.com	6	TLSv1.2	No (but handshakes did fail)	-1	-1	-	unknown	passed	passed	passed	passed	
events.appsflyer.com	6	TLSv1.2	No	-1	-1	-	unknown	passed	passed	Nothing observed	passed	
app2.babylonpartners.com	6	TLSv1.2	No (but handshakes did fail)	-1	-1	-	unknown	failed	passed	Nothing observed	passed	
app.babylonpartners.com	6	TLSv1.2	No	-1	1	-	unknown	failed	passed	Nothing observed	passed	Possible leak (Body)
settings.crashlytics.com	5		Yes	-1	-1	-	unknown	Nothing observed	passed	Nothing observed	passed	
services.babylonpartners.com	5	TLSv1.2	No (but handshakes did fail)	-1	1	-	unknown	Nothing observed	passed	Nothing observed	passed	Possible leak (Body)
gsp10-ssl.ls.apple.com	2	TLSv1.2	No (but handshakes did fail)	-1	-1	-	unknown	passed	Nothing observed	Nothing observed	Nothing observed	
api-glb-drf.smoot.apple.com	2	TLSv1.2	No	-1	-1	-	unknown	passed	Nothing observed	Nothing observed	Nothing observed	

Results

App Selection

Apps are selected by popularity in a relevant category from the Apple App Store as well as from Google Play Store. As mHealth apps are tested, the *medical* category is the most relevant. To diversify the test pool as much as possible, lists of most downloaded, free apps from three countries are considered. European privacy regulations are part of the considerations in this paper, therefore Germany, France and the United Kingdom—the most populated countries in Europe—were selected. The top 10 lists have been retrieved from App Annie on January 10, 2017 [57]. Top lists for a specific day are available after registration on the website. Whereas more exact app descriptions and results are available in the Multimedia Appendices 2 and 3, the apps were further categorized for better understanding of the test results. The categories of the apps tested are summarized in Table 1. The categories for each of the apps are part of the app descriptions in Multimedia Appendix 3.

Performing the Test Using the BProxy Tool

The test results were obtained utilizing the BProxy tool. The tool displayed results on a per domain basis. The first step in analyzing the output of the tool is the filtering of domains belonging to the app. These domains appear on top of BProxy's output, as they are communicated more frequently. The second step is to differentiate between connections to servers belonging to an app and those belonging to analytics or advertising providers. Next, the results in the columns are considered. They can be interpreted directly and contribute to the results presented here. To be able to make assessments regarding the integrity of data displayed by an app and confidentiality between an app and its servers, BProxy displays all request and response pairs for every domain. Requests and responses with app servers are examined and evaluated regarding their impact on integrity and confidentiality. In some cases, further testing, such as modification of server responses to validate integrity concerns, was performed using the Charles Web Debugging Proxy Application [58].

Summarized Results

Detailed results can be found in the form of two tables for Android and iOS apps in Multimedia Appendix 2. These tables list the results for each characteristic separately for every app. Further details on the apps (developers, top 10 list positions, and short descriptions) can be found in Multimedia Appendix 3.

All tests have been performed between January 17, 2017 and January 27, 2017. The most recent versions of the apps have been downloaded from the respective stores shortly before testing. Table 2 shows the summarized results of our tests. As none of the tested apps facilitated certificate pinning, the row was therefore omitted from the table.

The table shows that there are slightly more security issues in apps on the iOS platform in our test pool.

Table 1. Assigned categories of the tested apps.

Assigned category	Android, n (N=25)	iOS, n (N=28)	Total, n (N=53)
Pregnancy or fertility related	8	13	21
Drug information	2	1	3
Reference or learning	5	3	8
Consulting or communication	5	6	11
Health and fitness	3	3	6
Others	2	2	4

Table 2. Summarized table of results for Android and iOS apps.

Security issues	Android, n	iOS, n	Total, n
1. Servers outside European Union countries	7	8	15
2. No transport layer security for connections	7	12	19
3. Cookies or secure tokens send over insecure connections	4	7	11
4. Integrity of content displayed in the app compromised	8	13	21
5. Username and password sent over insecure connections	1	2	3
6. Confidentiality between user and app provider compromised	3	5	8
7. Certificate validation issues present	1	1	2

The most consequential issue observed is the omission of any kind of TLS (No transport layer security for connections) for connections present in 19 apps (36%). Insecure connections can lead to integrity (Integrity of content displayed in the app compromised) and confidentiality (Confidentiality between user and app provider compromised) breaches, as well as to exposed cookies, tokens (Cookies or secure tokens send over insecure connections), and user credentials (Username and password

sent over insecure connections). The semantic here was that as soon as a single unencrypted connection was used, the app was counted as not using TLS. Although the other issues are considered separately, they are more likely to occur in apps that fail to apply TLS for server connections.

Apps that do use TLS-secured connections were tested regarding their certificate validation mechanism as described. A failure to validate a certificate correctly (Certificate validation issues present) can expose all traffic sent through the TLS-secured connection to be exposed. This renders integrity, confidentiality, and authenticity protections otherwise offered by TLS useless. Two apps (4%) failed to validate server certificates correctly.

A total of 21 apps (40%) failed to protect the integrity of data they display, and a total of 11 apps (21%) failed to protect session data in transport (cookies or tokens), thus enabling attackers to hijack a session. Three mHealth apps (6%) sent user log-in credentials over insecure connections, whereas 8 (15%) compromised confidentiality of communication between the app and its servers.

Additionally, 15 apps (28%) used servers outside the European Union (EU). However, 31 more apps (58%) used analytics or advertising services outside EU countries, bringing the number of apps that communicated with servers outside the EU to 46 (87%).

In the most severe cases, apps transmit data (health data, usernames, and passwords) completely unprotected (ie, *iCare Health Monitor*). Other apps fetch menu structure and update prompt semantics (when and what to display when the app should be updated) through insecure channels. This enables third parties in privileged positions to hijack vulnerable parts of the app. Confidentiality issues were also popular, mostly because of the use of unsecured connections to retrieve content specific to a user's interest or condition. The *Pregnancy+* app, for example, automatically retrieves data through an unprotected (HTTP) connection for the week of the user's pregnancy. This can expose the state of the pregnancy to a third party.

Most popular analytics providers used up-to-date transport security standards. There was no general difference between the security concerns found in iOS and Android apps. However, single apps that exist on both platforms do show different security characteristics. For example, whereas the iOS version of the *Pregnancy+ app* is using a secure connection for log-in and transmission of data, the Android app does not use any kind of transport security. Similarly, the iOS version of the *babylon health online doctor* fails one of the certificate validation tests for one specific domain. This issue does not exist in the Android version of the app.

An issue was discovered in the Android version of the German *Apotheke vor Ort* app. The app does use a secure connection but accepts any (even invalid) certificates from the server. This is very problematic, among other things, because the app offers the possibility to send prescriptions (listing diagnosis, treating practitioner, and other sensitive medical details) to local pharmacies.

Developers of apps with critical test results were informed about the issues found in their apps before publication of this paper. As of March 23, 2017, a total of 5 developers reacted to the information shared with them. Four of the answers received were constructive.

Discussion

It was found that out of 53 apps tested from the three European App Stores for Android and iOS, 21/53 (40%) showed critical results. Out of these 21 apps, all failed to guarantee the integrity of data displayed. A total of 18 apps leaked private data or were observable in a way that compromised confidentiality between apps and their servers; 17 apps used connections without any protection; and 2 apps failed to validate certificates correctly. None of the apps tested utilized certificate pinning. Many apps employ analytics or ad providers, thereby undermining user privacy.

Common Security Concerns

The results show the following:

1. Analytics services are almost universally used in the apps under testing. Medical apps often handle sensitive data. Analytics services collect data without consideration of the kind of app using the software. Not only do these providers collect data that should potentially be protected, they also often are located outside EU countries and therefore not bound by EU regulations.
2. Many apps tested still use insecure endpoints or a mix of secure and insecure ones (19/53, 36%). Medical and health-related apps require protection of patient's data (authenticity of the apps communication partner and confidentiality between patient and app) and should display uncorrupted data (integrity). The lack of any kind of connection security results in the most severe security risks for users, patients, and providers of mHealth apps.
3. In this paper, pinning in any form was nonexistent in the tests. Whereas a crash reporting and analytics provider and Apple client software utilized pinning during the iOS tests, none of the apps under test on either platform utilized the technique for all connections to their servers.
4. Certificate validation seems to work fine for most apps that use secure connections (35/37, 95%); this is most likely because higher level programming interfaces are used. There are, however, cases in which apps accepted untrustworthy certificates. In one case, the app *Apotheke vor Ort* accepted all certificates, rendering TLS essentially useless. This is dangerous, as it seems to use a secure connection until certificate-focused tests are performed. In another problematic case, only one test failed for one domain, indicating an implementation error inside a library used by the app *babylon health online doctor*.

Comparison With Prior Work

The results presented in this paper were gathered by in-depth inspection and evaluation of the network traffic of selected mHealth apps. This differentiates the approach from metadata-based analysis on a fundamental level [26,27]. On a technical level, it is more comparable to Gagnon et al's AndroSSL analysis [29]. What makes the approach presented here different from that of AndroSSL is that AndroSSL must

be locally run, whereas BProxy can be run as a Web service and used by third parties to test their apps. AndroSSL is limited to the Android platform, as it relies on apps running in an emulator. Whereas iOS apps can be tested in a simulator during development, it is not possible to execute iOS binaries from the App Store on the iOS simulator. They are built for the Advanced RISC Machine architecture while the simulator requires them to be built for the Intel x86_64 architecture. The research presented in the AndroSSL publication was not aimed at mHealth-related apps, but only at transport security issues in Android apps.

Lastly, He et al's approach of in-depth analysis of apps is more manual than the approach presented here, but it does include more characteristics [28]. Although it does include log and storage analysis, the analysis of transport security is limited to detection of completely unencrypted traffic. For example, old TLS versions or certificate validation issues are not considered.

Recommended Practices for Transport Security

Most serious security issues are a result of missing or inconsistently implemented security measures. The recommended practices that help mitigate the security issues found are as follows:

- Use HTTPS calls exclusively. Be sure to keep the server (and the client) up to date. Enforce the most current TLS version and prevent a fallback to anything older than TLS 1.2 for any connections.
- Pinning is an option. As any certificate will expire, public key pinning or pinning the certificates of a smaller set of CAs can be a workable alternative to pinning to one certificate exclusively.

Up-to-date TLS for every connection from a mobile app prevents most security issues for data during transport. The following points represent suggestions when a full transition to this is unwanted or impossible for some reason:

- Only send usernames and passwords through secure connections.
- Session cookies or authorization tokens should not be sent over insecure connections.
- Loading resources over an insecure connection can leak user activity to an interested third party in a privileged position. Use secure connections to prevent this.

Because this paper focused on European mHealth apps, the location of servers should be kept in mind. Using a server in the EU gives the data on these servers special protection under European privacy regulations [10].

As mHealth apps often handle sensitive patient data, the use of third-party advertising and mobile analytics services should be seriously questioned and, if possible, avoided, or an opt-out option should be offered [11,59]. Analytics providers collect data not only to present it to the app developer but often also to mine information. The same caution should be exercised when considering the use of advertising services. They also enable extensive user tracking and thus pose a confidentiality risk [59]. Most third-party advertising and analytics services are based outside EU borders and legislation.

Some third-party services offer to deliver updates to apps via untrustworthy and unofficial update channels (not through Google's Play Store or Apple's App Store). The security implications of the use of these services are far-reaching and potentially open apps up to remote code injection, putting users at risk of confidentiality breaches and invalidating app integrity [60,61]. Third-party frameworks that use this technique should also be avoided.

Conclusions

The tests show that many mHealth apps do not apply sufficient transport security measures. The most common security issue was the use of any kind of unprotected connection. Some apps used secure connections only for selected tasks, leaving all other traffic vulnerable.

Authors' Contributions

JM implemented the BProxy tool, performed the tests on the apps, analyzed the results, and drafted the paper. CMF and TJ contributed to the experimental design and the analysis of the results and critically revised the paper. All authors approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Architectural diagram of the BProxy tool.

[PNG File, 182KB - mhealth_v5i10e147_app1.png]

Multimedia Appendix 2

Tables containing detailed results of every app under testing.

[PDF File (Adobe PDF File), 73KB - mhealth_v5i10e147_app2.pdf]

Multimedia Appendix 3

Tables listing the apps tested, their top list positions, developer, and a short description.

[PDF File (Adobe PDF File), 90KB - mhealth_v5i10e147_app3.pdf]

References

1. Adibi S, editor. Mobile Health: A Technology Road Map. Cham, Switzerland: Springer International Publishing; 2015.
2. Istepanian RS, Laxminarayan S, Pattichis CS, editors. M-Health: Emerging Mobile Health Systems. New York: Springer; 2006.
3. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? J Am Med Assoc 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
4. Istepanian R, Jovanov E, Zhang YT. Introduction to the special section on m-Health: beyond seamless mobility and global wireless health-care connectivity. IEEE Trans Inf Technol Biomed 2004 Dec;8(4):405-414. [Medline: 15615031]
5. Sunyaev A. Consumer facing health care systems. e Serv J 2014 Jan;9(2):1-23. [doi: 10.2979/eservicej.9.2.1]
6. Mechael PN. The case for mHealth in developing countries. Innov Technol Gov Global 2009 Jan;4(1):103-118. [doi: 10.1162/itgg.2009.4.1.103]
7. Puspitasari L, Ishii K. Digital divides and mobile Internet in Indonesia: impact of smartphones. Telemat Inform 2016 May;33(2):472-483. [doi: 10.1016/j.tele.2015.11.001]
8. 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]
9. Han D, Zhang C, Fan X, Hindle A, Wong K, Stroulia E. Understanding Android fragmentation with topic analysis of vendor-specific bugs. 2012 Presented at: 19th Working Conference on Reverse Engineering; October 15-18, 2012; Kingston, ON, Canada p. 83-92. [doi: 10.1109/WCRE.2012.18]
10. Eur-lex.europa. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC URL: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=DE [WebCite Cache ID 6oF8KwRGk]
11. European Commission. Ec.europa. 2016. Privacy code of conduct on mobile health apps URL: https://ec.europa.eu/digital-single-market/en/privacy-code-conduct-mobile-health-apps [accessed 2017-03-30] [WebCite Cache ID 6pLXQOiHf]
12. Simon SR, Evans JS, Benjamin A, Delano D, Bates DW. Patients' attitudes toward electronic health information exchange: qualitative study. J Med Internet Res 2009;11(3):e30 [FREE Full text] [doi: 10.2196/jmir.1164] [Medline: 19674960]
13. Dhopeshwarkar RV, Kern LM, O'Donnell HC, Edwards AM, Kaushal R. Health care consumers' preferences around health information exchange. Ann Fam Med 2012;10(5):428-434 [FREE Full text] [doi: 10.1370/afm.1396] [Medline: 22966106]
14. Möller B, Duong T, Kotowicz K. This POODLE bites: exploiting the SSL 3.0 fallback. Secur Advis 2014:1-4 [FREE Full text] [doi: 10.1145/1629575.1629606]
15. Carvalho M, DeMott J, Ford R, Wheeler DA. Heartbleed 101. IEEE Secur Priv 2014 Jul;12(4):63-67. [doi: 10.1109/MSP.2014.66]
16. Mehta N. Heartbleed. 2014. The Heartbleed Bug CVE-2014-0160 URL: http://heartbleed.com [accessed 2017-03-10] [WebCite Cache ID 6orAuJ8vf]
17. Sheffer Y, Holz R, Saint-Andre P. Tools.ietf. 2015. Summarizing known attacks on transport layer security (TLS) and datagram TLS (DTLS) URL: https://tools.ietf.org/html/rfc7457 [accessed 2017-09-18] [WebCite Cache ID 6tZNPoQ5k]
18. Louridas P. Static code analysis. IEEE Softw 2006 Jul;23(4):58-61. [doi: 10.1109/MS.2006.114]
19. Novak J, Krajnc A, Zontar R. Taxonomy of static code analysis tools. 2010 Presented at: 33rd International Convention MIPRO; May 24-28, 2010; Opatija, Croatia p. 418-422.
20. Chess B, McGraw G. Static analysis for security. IEEE Secur Priv 2004 Nov;2(6):76-79. [doi: 10.1109/MSP.2004.111]
21. Strafach W. Medium. 2017. 76 Popular apps confirmed vulnerable to silent interception of TLS-protected data URL: https://medium.com/@chronic_9612/76-popular-apps-confirmed-vulnerable-to-silent-interception-of-tls-protected-data-2c9a2409dd1 [accessed 2017-02-09] [WebCite Cache ID 6o9IzSitX]
22. Onwuzurike L, De Cristofaro E. Danger is my middle name: experimenting with SSL vulnerabilities in Android apps. New York, USA: ACM; 2015 Presented at: 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks - WiSec 2015; June 22-26, 2015; New York, NY. [doi: 10.1145/2766498.2766522]
23. Mylonas A, Meletiadis V, Tsoumas B, Mitrou L, Gritzalis D. Smartphone forensics: a proactive investigation scheme for evidence acquisition. In: Information Security and Privacy Research. 2012 Presented at: 27th IFIP TC 11 Information Security and Privacy Conference, SEC 2012; June 4-6, 2012; Heraklion, Crete, Greece. [doi: 10.1007/978-3-642-30436-1_21]
24. OWASP. 2016. Mobile top 10 2016 URL: https://www.owasp.org/index.php/Mobile_Top_10_2016-Top_10 [accessed 2017-02-09] [WebCite Cache ID 6o9JIoR8V]
25. Fahl S, Harbach M, Muders T, Baumgärtner L, Freisleben B, Smith M. Why Eve and Mallory love Android: an analysis of Android SSL (in)Security. USA: ACM; 2012 Presented at: ACM Conference on Computer Communications Security; October 16-18, 2012; New York, NY p. 50-61. [doi: 10.1145/2382196.2382205]

26. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
27. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The top Chinese mobile health apps: a systematic investigation. J Med Internet Res 2016 Aug 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
28. 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]
29. Gagnon F, Ferland MA, Fortier MA, Desloges S, Ouellet J, Boileau C. AndroSSL: A platform to test Android applications connection security. In: Foundations and Practice of Security. Cham, Switzerland: Springer International Publishing; 2015:294-302.
30. Gagnon F, Lafrance F, Frenette S, Hallé S. AVP - an Android virtual playground. Portugal: SCITEPRESS - Science and and Technology Publications; 2014 Presented at: 5th International Conference on Data Communication Networking; August 28-30, 2014; Vienna, Austria p. 13-20. [doi: 10.5220/0005016500130020]
31. Georgiev M, Iyengar S, Jana S, Anubhai R, Boneh D, Shmatikov V. The most dangerous code in the world: validating SSL certificates in non-browser software. New York, USA: ACM; 2012 Presented at: ACM Conference on Computer Communications Security - CCS; October 16-18, 2012; Raleigh, NC p. 38-49. [doi: 10.1145/2382196.2382204]
32. Fielding R, Gettys J, Mogul J, Frystyk H, Masinter L, Leach P, et al. Tools.ietf. 1999. Hypertext Transfer Protocol - HTTP/1.1 URL: https://tools.ietf.org/html/rfc2616 [accessed 2017-09-18] [WebCite Cache ID 6tZREH58V]
33. Rescorla E. Tools.ietf. 2000. HTTP over TLS URL: http://tools.ietf.org/html/rfc2818 [accessed 2017-09-18] [WebCite Cache ID 6tZRpcpC7]
34. Zimmermann H. OSI reference model—the ISO model of architecture for open systems interconnection. IEEE Trans Commun 1980 Apr;28(4):425-432. [doi: 10.1109/TCOM.1980.1094702]
35. Maurer U. Modelling a public-key infrastructure. UK: Springer-Verlag; 1996 Presented at: 4th European Symposium on Research in Computer Security: Computer Security; September 25-27, 1996; London, UK p. 325-350.
36. Housley R, Ford W, Polk W, Solo D. Tools.ietf. 1999. Internet X.509 public key infrastructure certificate and CRL profile URL: https://tools.ietf.org/html/rfc2459 [accessed 2017-09-18] [WebCite Cache ID 6tZSApaoX]
37. Meyer C, Schwenk J. Lessons learned from previous SSL/TLS attacks-a brief chronology of attacks and weaknesses. In: Kim Y, Lee H, Perrig A, editors. Information Security Applications. WISA 2013. Lecture Notes in Computer Science, Volume 8267. Cham, Switzerland: Springer International Publishing; 2014.
38. Rizzo J, Duong T. Practical padding oracle attacks. 2010 Presented at: 4th USENIX Conference on Offensive Technologies; 2010; Washington, DC.
39. Brubaker C, Jana S, Ray B, Khurshid S, Shmatikov V. Using frankencerts for automated adversarial testing of certificate validation in SSL/TLS implementations. 2014 Presented at: IEEE Symposium on Security and Privacy; May 18-21, 2014; San Jose, CA p. 114-129. [doi: 10.1109/SP.2014.15]
40. Durumeric Z, Ma Z, Springall D, Barnes R, Sullivan N, Bursztein E, et al. The security impact of HTTPS interception. 2017 Presented at: Network and Distributed Systems Symposium (NDSS 2017); 2017; San Diego, CA. [doi: 10.14722/ndss.2017.23456]
41. Brubaker C. Android-developers. 2016. Changes to trusted certificate authorities in Android Nougat URL: https://android-developers.googleblog.com/2016/07/changes-to-trusted-certificate.html [accessed 2017-02-09] [WebCite Cache ID 6o9JsGPTS]
42. Faruki P, Bharmal A, Laxmi V, Ganmoor V, Gaur MS, Conti M, et al. Android security: a survey of issues, malware penetration, and defenses. IEEE Commun Surv Tutor 2015;17(2):998-1022. [doi: 10.1109/COMST.2014.2386139]
43. Joshi J, Parekh C. Android smartphone vulnerabilities: a survey. 2016 Presented at: International Conference on Advances in Computing, Communication, & Automation (ICACCA); April 8-9, 2016; Dehradun, India.
44. Statista. 2016. Distribution of Android operating systems used by Android phone owners in September 2017, by platform versions URL: https://www.statista.com/statistics/271774/share-of-android-platforms-on-mobile-devices-with-android-os/ [accessed 2017-03-01] [WebCite Cache ID 6odWECn6x]
45. Apple. 2016. iOS security URL: https://www.apple.com/business/docs/iOS_Security_Guide.pdf [accessed 2017-02-13] [WebCite Cache ID 6oFHHo9Cw]
46. Elenkov N. Android Security Internals: An In-Depth Guide to Android's Security Architecture. San Francisco, CA: No Starch Press; 2014.
47. Evans C, Palmer C, Sleevi R. Tools.ietf. 2015. Public key pinning extension for HTTP URL: https://tools.ietf.org/html/rfc7469 [accessed 2017-09-18] [WebCite Cache ID 6tZTmFcTz]
48. Chloe. Labs.detectify. 2016. What HPKP is but isn't URL: https://labs.detectify.com/2016/07/05/what-hpkp-is-but-isnt/ [accessed 2017-02-13] [WebCite Cache ID 6oFFFMLeD]
49. Kristol D, Montulli L. Ietf. 1997. HTTP state management mechanism URL: https://www.ietf.org/rfc/rfc2109.txt [accessed 2017-09-18] [WebCite Cache ID 6tZUQTdMf]

50. Sivakorn S, Polakis I, Keromytis AD. The cracked cookie jar: HTTP cookie hijacking and the exposure of private information. 2016 Presented at: IEEE Symposium on Security and Privacy (SP); May 22-26, 2016; San Jose, CA p. 724-742. [doi: 10.1109/SP.2016.49]
51. Liu AX, Kovacs JM, Gouda MG. A secure cookie scheme. Comput Netw 2012 Apr;56(6):1723-1730. [doi: 10.1016/j.comnet.2012.01.013]
52. Franks J, Hallam-Baker P, Hostetler J, Lawrence S, Leach P, Luotonen A, et al. Tools.ietf. 1999. HTTP authentication: basic and digest access authentication URL: https://tools.ietf.org/html/rfc2617 [accessed 2017-08-14] [WebCite Cache ID 6si5gpzVj]
53. Jones M, Hardt D. Tools.ietf. 2012. The OAuth 2.0 authorization framework: bearer token usage URL: https://tools.ietf.org/html/rfc6750 [accessed 2017-08-14] [WebCite Cache ID 6si5ZZo1Z]
54. OWASP. 2016. OWASP zed attack proxy project URL: https://www.owasp.org/index.php/OWASP_Zed_Attack_Proxy_Project [accessed 2017-03-10] [WebCite Cache ID 6orKCFlOM]
55. Fielding RT. ics.uci. Irvine, CA: University of California; 2000. Architectural Styles and the Design of Network-based Software Architectures URL: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation_2up.pdf
56. Github. 2017. BProxy - Client-focused transport security testing proxy URL: https://github.com/j4nnis/bproxy [accessed 2017-08-14] [WebCite Cache ID 6si5p0pJU]
57. Appannie. 2017. Accelerate success for your app business URL: https://www.appannie.com/ [accessed 2017-09-18] [WebCite Cache ID 6tZVUhaWZ]
58. Charlesproxy. 2017. Charles - Web debugging proxy application URL: https://www.charlesproxy.com [accessed 2017-06-08] [WebCite Cache ID 6r3zAhYoP]
59. Chen T, Ullah I, Kaafar MA, Boreli R. Information leakage through mobile analytics services. 2014 Presented at: 15th ACM International Workshop on Mobile Computing and Applications - HotMobile'14; 2014; Santa Barbara, CA p. 1-6. [doi: 10.1145/2565585.2565593]
60. Castelluccia C, Kaafar MA, Tran MD. Betrayed by your ads! Reconstructing user profiles from targeted ads. 2012 Presented at: 12th International Conference on Privacy Enhancing Technologies (PETS); July 11-13, 2012; Vigo, Spain p. 1-17. [doi: 10.1007/978-3-642-31680-7_1]
61. Jing X, Zhaofeng C, Jimmy S. Fireeye. 2016. Hot or not? The benefits and risks of iOS remote hot patching URL: https://www.fireeye.com/blog/threat-research/2016/01/hot_or_not_the_bene.html [accessed 2017-03-09] [WebCite Cache ID 6opguymJJ]

Abbreviations

CA: certificate authority
EU: European Union
HPKP: HTTP public key pinning
HTTP: Hypertext Transfer Protocol
HTTPS: Hypertext Transfer Protocol Secure
mHealth: mobile health
MitM: man-in-the-middle
REST: representational state transfer
SSL: secure socket layer
TLS: transport layer security

Edited by D Zhong; submitted 30.03.17; peer-reviewed by F Gagnon, T Dehling, M Hölbl, A Kamisalic; comments to author 18.05.17; revised version received 11.07.17; accepted 15.08.17; published 18.10.17.

Please cite as:
Müthing J, Jäschke T, Friedrich CM
Client-Focused Security Assessment of mHealth Apps and Recommended Practices to Prevent or Mitigate Transport Security Issues
JMIR Mhealth Uhealth 2017;5(10):e147
URL: https://mhealth.jmir.org/2017/10/e147/
doi:10.2196/mhealth.7791
PMID:29046271

Original Paper

Attention Deficit Hyperactivity Disorder: Is There an App for That? Suitability Assessment of Apps for Children and Young People With ADHD

Lauren Powell[1], BSc (Hons), MSc; Jack Parker[1], BSc (Hons), PhD; Naomi Robertson[2]; Valerie Harpin[3], MBBChir, MD, FRCP, FRCPCH

[1]School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

[2]Department of Psychology, University of Sheffield, Sheffield, United Kingdom

[3]Ryegate Children's Centre, Sheffield Children's NHS Foundation Trust, Sheffield, United Kingdom

Corresponding Author:
Lauren Powell, BSc (Hons), MSc
School of Health and Related Research
University of Sheffield
Regent Court, 30 Regent Street
Sheffield, S1 4DA
United Kingdom
Phone: 44 11422 ext 28275
Fax: 44 1142220749
Email: l.a.powell@sheffield.ac.uk

Abstract

Background: Attention-deficit/hyperactivity disorder (ADHD) is a complex highly comorbid disorder, which can have a huge impact on those with ADHD, their family, and the community around them. ADHD is currently managed using pharmacological and nonpharmacological interventions. However, with advances in technology and an increase in the use of mobile apps, managing ADHD can be augmented using apps specifically designed for this population. However, little is known regarding the suitability and usability of currently available apps.

Objective: The aim of this study was to explore the suitability of the top 10 listed apps for children and young people with ADHD and clinicians who work with them. It is hypothesized that mobile apps designed for this population could be more suitably designed for this population.

Methods: The top 10 listed apps that are specifically targeted toward children and young people with ADHD in the United Kingdom were identified via the Google Play (n=5) and iTunes store (n=5). Interviews were then undertaken with 5 clinicians who specialize in treating this population and 5 children and young people with ADHD themselves, to explore their opinions of the 10 apps identified and what they believe the key components are for apps to be suitable for this population.

Results: Five themes emerged from clinician and young people interviews: the accessibility of the technology, the importance of relating to apps, addressing ADHD symptoms and related difficulties, age appropriateness, and app interaction. Three additional themes emerged from the clinician interviews alone: monitoring symptoms, side effects and app effect on relationships, and the impact of common comorbid conditions. The characteristics of the apps did not appear to match well with the views of our sample.

Conclusions: These findings suggest that the apps may not be suitable in meeting the complex needs associated with this condition. Further research is required to explore the value of apps for children and young people with ADHD and their families and, in particular, any positive role for apps in the management of ADHD in this age group. A systematic review on how technology can be used to engage this population and how it can be used to help them would be a useful way forward. This could be the platform to begin exploring the use of apps further.

(JMIR Mhealth Uhealth 2017;5(10):e145) doi:10.2196/mhealth.7371

KEYWORDS

attention deficit disorder with hyperactivity; mobile applications; technology

Introduction

ADHD, Technology, and Mobile Apps

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder characterized by three core symptoms: inattention, impulsivity, and hyperactivity, which can have a profound impact on the individual, their family, and their community [1-3]. It is a highly comorbid [4-6] chronic disorder and has a prevalence of 3% to 5% in school-age children worldwide [7]. Furthermore, 80% to 85% of these children will continue to be impaired by their ADHD symptoms as adolescents and 60% as adults [2,8-16]. Indeed, the presence of ADHD increases the risk of premature death [17]. Those whose ADHD persists into adulthood are more likely to engage in criminality and substance abuse [11,18,19]. Globally, ADHD management involves a combination of nonpharmacological and pharmacological interventions [20-22]. In mild to moderate cases, behavioral interventions such as psychoeducation and cognitive behavioral therapy are used alone, whereas in more severe cases, it is recommended that both pharmacological and nonpharmacological approaches are used concurrently [20,21]. Currently, the resources available to the service provider may limit interventions. Therefore, interventions that can offer support with minimal input from the clinic or school and can be generalizable would be highly desirable. Contemporary forms of engaging children and young people, such as the use of technology, could have potential in facilitating greater self-awareness, improving self-management skills or management for carers, and managing the condition into adulthood [23].

Xu et al reviewed 19 studies assessing technology use in students with ADHD in the age group of 4 to 19 years. The authors concluded that as yet there is very little evidence to support the effectiveness of such interventions [24]. Another recent review of what the authors describe as the most representative studies of the past decade assessed the findings of research that investigated the use of varying technologies with young people diagnosed with ADHD [25]. Studies included involved the use of a handheld device to help organize daily activities [26] or self-monitor symptoms [27], software to improve reading speed [28], and games to improve mathematical ability [29]. The success of these technologies was measured in a number of ways, including observational data [26-29], the Behavioral Assessment of Dysexecutive syndrome [26], qualitative interviews [26], reading speed, and time to complete assignments [28]. The authors concluded that they believe technology can enhance the learning of people with attention difficulties; however, the evidence base for this remains limited.

Another study looked at a computer mission game that aims to promote behavioral learning and organization of daily skills such as time management and planning or organizing [30]. Children played the game either 3 or 8 times a fortnight. A user satisfaction survey showed that between-group differences for game satisfaction were not observed, but children did enjoy the game and reported learning from it [30].

Technology has also been used by clinicians to aid diagnosis of ADHD and to monitor outcomes. For example, the quantified behavior (Qb) Test uses the Continuous Performance Test to produce a visual graph of the three core ADHD symptoms [31] and has been used to assist with the diagnosis of ADHD [32]. The primary outcomes of this study are time to diagnosis and diagnosis accuracy. The secondary outcome measures are clinician's diagnostic confidence and routine clinical outcome measures. The authors are also conducting a qualitative assessment of the feasibility and acceptability of incorporating the QbTest into routine practice. A Web-based technology, *Health Tracker*, has also been used for parents, children, and professionals to track the long-term outcomes of children and young people with ADHD to enable more effective treatments and a more efficient service delivery [33].

The use of mobile phones and mobile devices has risen dramatically over recent years. Indeed, a UK report cofunded by the European Union found that 93% of children and adolescents aged 9 to 16 years in the United Kingdom access the Internet at least weekly and over half of these do so via mobile devices such as mobile phones [34]. As a result, mobile apps are also increasingly popular, and there has been some success in using them to engage children or young people with ADHD. These apps can include games, information about ADHD diagnosis and ADHD treatment, various ADHD tests, task managements, and reminders [35]. These attempts often incorporate reward [36-38], bright colors, and varied visual stimuli [37,39-43]. Examples include using apps said to monitor behavior in ADHD [41], improve behavior [44], improve organization skills [39,43] address medication compliance [36,45], improve reading motivation and summarization [40], and improve cognition through the use of games [37].

Currently available research evaluating apps for children and young people with ADHD comprises single case studies [37], technology development reports [36,42], and small sample sizes [36,37,40,42-44]. The conclusions drawn by these evaluation attempts of these apps include benefits for children and young people with ADHD, such as the app can improve organization and time management, reduce conflicts with parents during morning routines [43], and improve academic improvement [37] and *ontask* behaviors [44]. These claims are based on small sample sizes of 2, 1, and 8, respectively. Each evaluation method varied, and there was little consistency in the way apps were assessed.

Although an increasing number of apps are being promoted for use by or with [46] individuals with ADHD, there is still little guidance to support the reliability, validity [35], and suitability of currently available apps. There are few rules and regulations around what apps are suitable and for whom. In England, a review of the National Health Service (NHS) Choices apps library was principally focused on three components: (1) compliance with the Data Protection Act, (2) evidence of efficacy, and (3) relevance to British individuals. Many identified apps did not have an evidence base, and it was shown that privacy and data security was not suitable [46]. Within a week, the library was gone.

The concept of the use of apps to manage ADHD in children and young people is in its infancy. The research base is thin and is hugely outweighed by the number of apps available, thus

suggesting that available apps are often not evidenced-based [35] and may not be suitable for their complex target population. Therefore, this study aims to identify and evaluate currently available apps that are aimed at this population by gaining the opinions of children and young people with ADHD and specialist clinicians. These opinions will be used to ascertain what they believe would make an app suitable for this population. It is hypothesized that the selected apps will not be suitable for this population, as apps generally have a thin or no evidence base, and this complex population has very specific needs.

Methodology

This research involved initially the identification of the top 10 listed apps aimed at children and young people diagnosed with ADHD. The apps were identified as top 10, which was deemed as a reasonable and manageable figure [47,48] as there is a limited number of apps a consumer will search for. The authors believe that the top 10 apps will be viewed as the best 10 apps as they are the top 10 apps for a reason. Subsequently, young people with ADHD tested the apps, and they were then interviewed to ascertain their views on the apps and to explore what they believed the key components are for apps to be helpful for them. Clinicians were also interviewed to explore their insights into how to make apps successful for this population.

Research Question

Are the top 10 listed apps specifically designed and marketed for children and young people with ADHD suitable, and what are the key components for apps to be suitable for this population?

Methods

Search and Identification of Mobile Apps

In June 2016, a search of mobile apps in the Apple iTunes store and the Android Google Play store in the United Kingdom was conducted. iTunes and Google Play were chosen, as they are the two largest and most popular app stores [49]. Apps for iPads and tablets were reviewed rather than phones, as apps are available on phone and tablet devices. For this study, apps available on tablets have been reviewed as they are easier to discuss with children, young people, and clinicians. These databases were selected as they display systematically organized app rankings defined by algorithms unique to each app store, commonly known as app store optimization (ASO). For Apple, the primary factor is the number of downloads; however, there are also many other secondary factors such as keywords and visuals [50]. Similarly, the Android database is filtered according to multiple criteria, including the volume of ratings, value of ratings, and download growth [51]. Although this gives rise to potential bias, as apps are selected according to the database's own ASO, it is to some extent unavoidable unless all of the search results are downloaded for testing [48]. Where the authors do acknowledge that other app stores such as Amazon, Windows, and Blackberry do exist, currently, these do not have enough of a market share to be considered.

The search term we used was "ADHD." This is because other search terms such as "ADD," "children," and "Young people" did not provide different results to the "ADHD" searches. The term "ADHD" was searched in both of the listed app stores.

Preliminary screening was conducted based on app titles, full marketing description, and screenshots of the apps potentially relevant for inclusion. The first five apps that fit the inclusion criteria (Textbox 1) were included from each app store, giving a total of 10 apps for inclusion.

Duplicate apps were then removed (see Figure 1). This was applicable if there was more than one version of an app. It was decided that the app version to be included was the app that appeared first on the app store list.

The five apps from iTunes that fit the inclusion criteria were downloaded onto an Apple iPad mini (model: A1489), and the five apps from Google Play were downloaded to a Samsung device (model: GT-P5220). The app contents were summarized by 2 members of the team (LP and NR). The apps were simply summarized into a tabular format to help assist clinician participants during the semistructured interviews (see Multimedia Appendix 1). Tables 1 and 2 also give a brief overview of the apps claims. Multimedia Appendix 1 was given to clinicians during interviews for their information.

Textbox 1. Inclusion and exclusion criteria of apps.

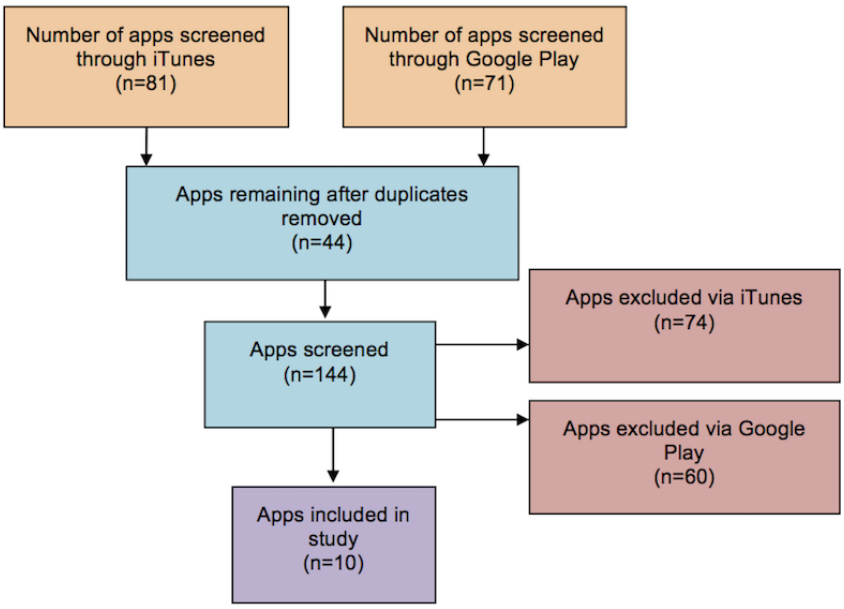
Inclusion criteria

- States aimed at attention-deficit/hyperactivity disorder (ADHD)
- The user is a child or young person with ADHD or attention deficit disorder (ADD)
- Mobile app
- App is available in the English language

Exclusion criteria

- Does not state that app is aimed at ADHD or ADD
- Not targeted at the child or young person with ADHD or ADD (eg, targeted at parents or clinicians only)
- Not a mobile app
- Not available in the English language
- Duplicate app

Figure 1. App selection process. The number of apps screened is the number of apps that had to be screened before identifying five apps for each database that fit the inclusion criteria. Eight apps were duplicated across both databases; therefore 144 apps were screened before the final ten were selected.



Participants and Recruitment

A convenience sample of 5 clinicians and 5 children and young people were invited to take part. Children and young people were recruited via a Family Action group (a national charity that provides support for families) and a Web-based ADHD community group. Clinicians were recruited via the NHS (Child and Adolescent Mental Health Service and a Pediatric Neurodisability Service in the South Yorkshire region). Four clinicians were medical staff and one a specialist nurse. Participants were recruited until data [52] were achieved. Eligibility criteria for clinicians were that they had to be employed by a service that treats children and young people with ADHD and to specialize in treating this population themselves. Eligibility criteria for the children and young people were (1) must have a confirmed ADHD diagnosis and (2) must be in the age group of 6 to 17 years.

Procedure

The study received ethical approval from the University of Sheffield's School of Health and Related Research Ethics Committee (references 007880, 010768).

Clinicians were approached via email. A University of Sheffield researcher visited a group for parents of children with ADHD. Parents are referred to this group by NHS clinicians when their child receives a diagnosis of ADHD. The researcher explained the study to the parents and the parents discussed it with their children at home. Age-appropriate information sheets were also provided for the parents to give to their children. If their child wanted to take part in the study, the parent either contacted the researcher on their behalf or, with permission the researcher, contacted the parent to discuss and arrange an interview appointment. Ten separate semistructured interviews took place in the Sheffield region and were held in locations convenient to the participants (either NHS clinic for clinicians or home for the young people) from June to July 2016 for the clinicians and from October to November 2016 for the young people. Interviews lasted up to 60 min with the clinicians and up to 45 minutes with the child or young people.

Before the interviews, the clinicians provided written informed consent. Children and young people provided written assent and their parents consented on their behalf.

Information about the study was sent to all participants at least 1 week before interviews took place. At the beginning of each interview, the study was explained to the participants and questions were answered.

Clinicians were presented with all 10 apps identified. As two of the 10 apps were aimed at a younger age group (3-7 years) and all children or young people interviewed were over the age of 7 years, they were given eight apps to review. Participants were given the opportunity to use the apps themselves during the interview. Interview discussions occurred after participants examined each app. These discussions were guided by an interview schedule covering their views in four key areas:

- What makes a successful app?
- What doesn't make a successful app?
- How could an app function benefit this population
- How could apps help manage ADHD or address difficulties in young people

The two groups provided two unique perspectives: a user perspective and a clinical perspective. Participants also completed a short questionnaire on their demographic characteristics.

Data Analysis

All interviews were audio-recorded and transcribed verbatim. Thematic analysis [53] was used to search for data patterns

within and across the participant groups. LP and JP independently identified codes and themes from the transcripts. Discrepancies were resolved through group discussion in an iterative fashion between authors. Themes identified aimed to capture the essence of the participants' views.

Additionally, participants identified characteristics they believe apps should include if they are to be suitable for this population. These characteristics are presented in Table 3 and are accounted for within the qualitative analysis. Authors tabulated these characteristics and assessed each of the 10 apps. They documented how many of these characteristics identified by participants were present in the apps assessed in this study.

Results

ADHD: Is There an App for That?

Five apps were identified from Google Play and five from iTunes in July 2016. Tables 1 and 2 describe the claims the apps make within their individual descriptions and their contents. Table 3 demonstrates the app characteristics identified by participants during semistructured interviews and which apps possessed these characteristics. These characteristics are accounted for within the themes that emerged from the qualitative analysis and were discussed and agreed by participants.

Table 1. Characteristics of the apps downloaded from iTunes.

App	App claims	How apps meet their claims
1	Improves attention, concentration, focus, perceptual reasoning, academic performance, and inhibition impairments	Games to improve cognition
2	Improves self-control, reduces hyperactivity, improves attention, concentration, and focus	Mindfulness training
3	Improves self-control, reduces hyperactivity, and improves attention, concentration, and focus	Mindfulness training
4	Improves attention, concentration, and focus	Different version of same game involves responding to stimuli as quickly as possible.
5	Improves academic performance	Games and lessons (for user to watch)

Table 2. Characteristics of the apps downloaded from Google Play apps.

App	App claims	How apps meet their claims
6	Addresses memory and provides information about ADHD[a]	Memory games, different levels, ADHD key concepts quiz. Dialogues with a cartoon character, links provided with ADHD information
7	Improves attention, concentration, and focus, and addresses memory	Three games (find all objects, find numbers, reaction times), 4 memory games
8	Visualize time moving	On-screen moving timer
9	Improves academic performance	Improve reading speed: different ways of presenting text, books for different reading abilities, comprehension quiz on text previously read
10	Provides motivation	Talking fitness avatar, games involving physical activity

[a]ADHD: attention-deficit/hyperactivity disorder.

Table 3. Summary of what makes an app suitable for a child or young person with attention-deficit/hyperactivity disorder (ADHD; left column), according to the children and young people with ADHD and the clinicians interviewed in this study and which apps identified in this study include these characteristics. Authors have examined each app against the criteria identified by participants and scored the apps out of 8 to highlight how they, mostly, are not in line with the needs of their target audience.

Characteristics identified by participants as likely to be positive	iTunes					Google Play				
	App 1	App 2	App 3	App 4	App 5	App 6	App 7	App 8	App 9	App 10
Visually pleasing (ie, includes bright colors)	Yes	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Allows personalization so user can relate to app	No	No	No	No	No	No	No	No	No	Yes, change avatar's clothes
Plays music	Yes	Yes, relaxation	Yes, relaxation	No	Briefly in some sections, not all	Yes	No	No	No	No
Provides audio feedback (ie, makes a sound when user interacts with app)	Yes	No	No	Yes, pig snorts at a correct response	Only in some sections	Yes	No	No	No	Yes
Involves instant reward	No	No	No	No	No	Yes, collect coins, different levels	No	No	No	Yes, user gains points
Is interactive	Yes, games	No	No	Yes	No	Yes	No	Yes	No	No
Involves symptom monitoring component or ADHD[a]-related monitoring component (eg, diet)	No	No	No	No	No	No	No	No	No	No
Involves component that encourages healthy relationships with others	No	No	No	No	No	No	No	No	No	No
App score (out of 8)	4	2	1	3	2	5	0	2	0	4

[a]ADHD: attention-deficit/hyperactivity disorder.

Table 4. Demographic characteristics of young people.

Unique ID	Age, in years	Gender	Length of ADHD[a] diagnosis (years, months)	Other diagnoses	Current prescribed ADHD medication	Medicated during interview?
YP1[b]	10	Male	1, 11	Autism spectrum disorder (ASD)	Concerta XL, melatonin	Yes
YP2	13	Female	0, 7	ASD, anxiety	Equasym XL	No
YP3	9	Male	6, 5	ASD	None	No
YP4	8	Male	0, 5	Not applicable	Equasym XL, Methylphenidate Immediate release	Yes
YP5	8	Female	1, 0	ASD, generalized anxiety disorder, sensory processing difficulties	Methylphenidate Immediate release	No

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]YP: young person.

Participant Characteristics

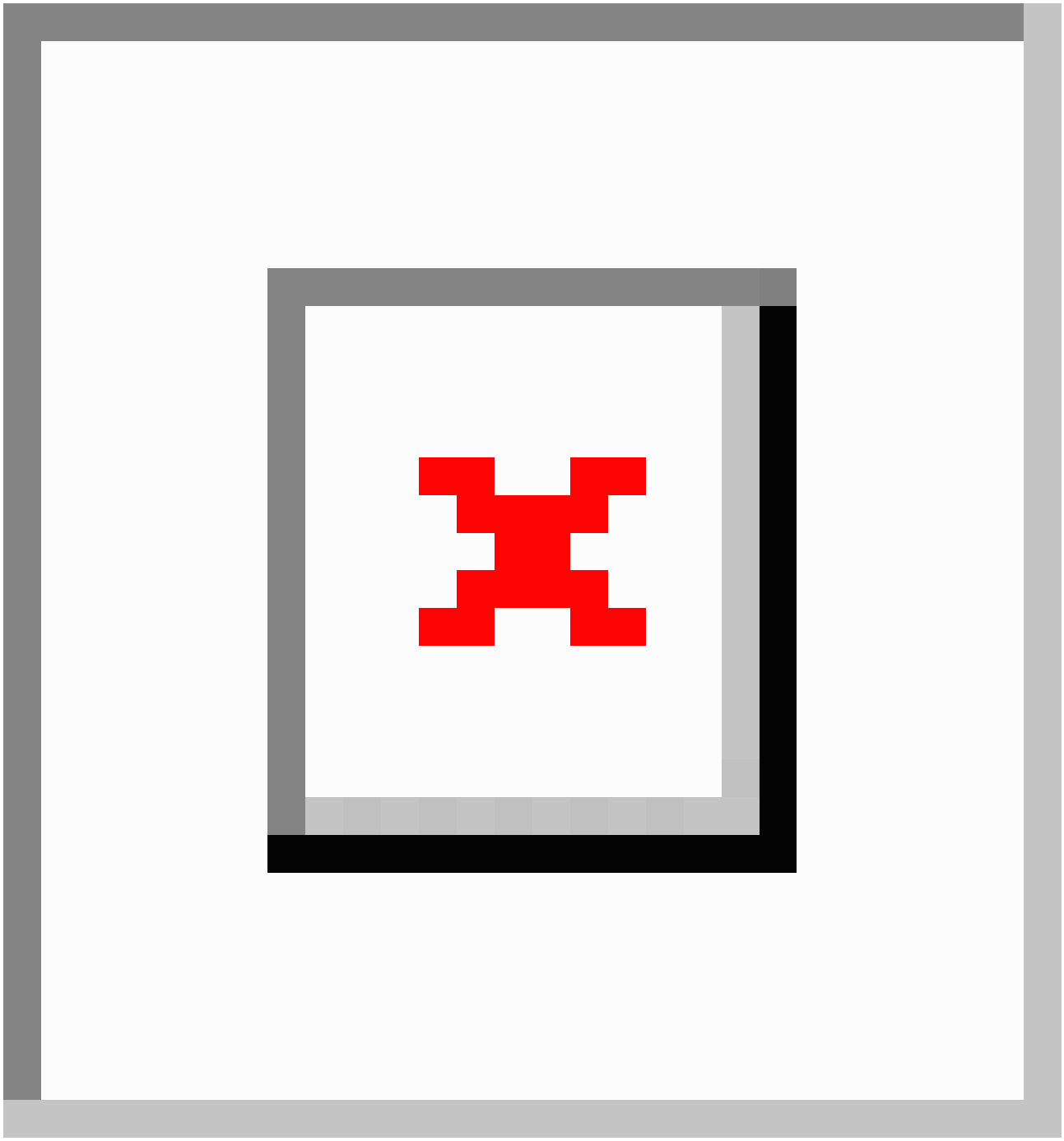
Ten participants, recruited in 2016, were included in this study: 5 clinicians and 5 young people diagnosed with ADHD aged 8 to 13 years (2 female, 3 males). Participant characteristics are reported in Table 4 (young people) and Table 5 (clinicians). Four of the five young people provided their ADHD medication as additional confirmation of diagnosis. One participant was not medicated. Four out of the five young people were recruited via an ADHD parent group in Sheffield. Parents were only referred to this group if their child was diagnosed with ADHD. Data were collected until data saturation was reached. Transcripts were available for all 10 interviews. During analysis, agreement between the 2 primary coders was high. Seven themes were identified in total (see Figure 2). Where similarities between children or young people and clinician views emerged, from the data, they are combined and discussed under the same themes. These themes are identified, compared, and discussed below with illustrative quotations.

Table 5. Demographic characteristics of clinicians demonstrating a total of more than 57 years of experience working with children and young people with attention-deficit/hyperactivity disorder (ADHD).

Unique ID	Gender	Length of time working with population (years)	Current job title
HCP1[a]	Male	4	Specialist registrar in Child and Adolescent Mental Health Service
HCP2	Female	11	Consultant child and adolescent psychiatrist
HCP3	Female	22	Community mental health nurse
HCP4	Female	15	Associate specialist in pediatric neurodisability
HCP5	Female	5	Consultant child and adolescent psychiatrist

[a]HCP: health care professional.

Figure 2. Summary of themes that emerged from the data and which participant perspectives they represent.



Eight themes emerged from the data and are presented below with supporting quotations. They cover views specifically expressed by clinicians, which include the following: that apps could be used to monitor symptoms of children and young people with ADHD, apps could have both a positive and a negative effect on the relationships of child or young person with ADHD with others, and the impact of common comorbid conditions. Other themes expressed the views of both the clinicians and the children or young people. These themes involve discussion around accessibility issues of the technology, how important it is that apps consider ADHD symptoms and related difficulties, they should be age appropriate, preferably have an element that can be personalized so that the user can relate to the app, and similarly, the user should be able to interact with the app rather than simply watching a video clip or listening to audio recordings.

Technology Accessibility

Young people (n=2) and clinicians (n=5) noted that the apps were not always reliable, as they often did not work properly or the log-in functions failed. This was a barrier when trying to engage them with the technology. It was recognized as “silly” (YP2) that the apps were all only available via iTunes or Google Play, which was problematic for 2 participants, as they did not have access to both. The young people also felt that it was a barrier to pay for apps, as they are not old enough to pay for them online (n=2). One participant noted that iPads could be

slow and not load quickly enough, which was considered as "boring" (YP2). One participant noted:

> *Some didn't register log-ins but that's just the people that made it messing up and that.* [YP1]

Another participated stated:

> *...you have to like buy them but that's annoying cos they should be free...I haven't even got a credit card.* [YP2]

Clinicians were often skeptical of the claims the app developers made (n=3) and believed they lacked underpinning evidence (n=3), as noted below:

> *...it's difficult through an app to build a child's kind of confidence.* [HCP1]
>
> *So they are trying to just say that going on an app makes you less impulsive and reduces your hyperactivity which I very very much doubt.* [HCP4]
>
> *...quite bold non evidenced based statements aren't they.* [HCP]

Importance of User Relating to the App

All 5 young people noted a number of points relating to their wish to relate to the app; audio was important to all participants. Two young people described one of the voices on the app as a "robot" (YP1 and YP3) and as having a "creepy voice" that goes "on and on and on" (YP1); they said they wanted the voice to have "more expression" (YP1). Young people and clinicians noted that the apps should be visually attractive (n=10) and "colorful" (YP2, YP3, and YP4) or "pretty" (YP5). They liked the idea of changing some characters in the apps to represent themselves (n=5) or somebody else such as Justin Bieber, their favorite pop icon (YP2). Similarly, many of the young people believed that the character in apps is important (n=2):

> *...cos if there were no sound you'd just play something with no sound and it'll be real boring.* [YP3]
>
> *...they (app developers) could change the voice a bit to like a character in like a like someone in a film.* [YP4]
>
> *Well his background needs changing to little boys bedroom...* [YP1]
>
> *...the trousers you could actually have trackies on like mine.* [YP4]

Similarly, the clinicians thought that the apps should be fun (n=3), visually attractive (n=5), use language (n=3) and accents (n=1) the child can relate to, and incorporate reward (n=1). It was discussed that relating to an app may not always be a positive experience for young people. One clinician stated that one of the apps appeared to be mimicking school, which a child may negatively relate to. Another clinician stated that a child might respond to a character on an app telling them what to do as opposed to a parent or school teacher:

> *...cos it just feels like being at school...and I think they've got to be fun.* [HCP3]
>
> *...the logo is very important...cos that's what teenagers will see...people with ADHD aren't going to read all the details they're going to be very impulsive aren't they and think oh that looks interesting and bang (download the app).* [HCP4]

Addressing Attention-Deficit/Hyperactivity Disorder (ADHD) Symptoms and Related Difficulties

Young people noted a number of issues that involve addressing their ADHD symptoms. For example, 2 participants liked the idea that apps could be used to relax them (ie, assist with their hyperactivity).

Participants noted that gaining an instant reward such as "leveling up" or gaining coins during a game made them feel "happy" (YP2). One participant found apps had the opposite effect when she was denied rewards, such as leveling up (YP5). Four participants said it is important that an app contains lots of variety to keep YPs attention. One participant stated:

> *...it [the app] should make it so every time you complete one [game level] you get little coins and you can buy backgrounds and that.* [YP1]

Another participant noted:

> *It makes you focus on the pattern so you relax.* [YP2]

Similarly, clinicians also believed that reward and variety in apps are important to engage the user. One clinician believed that apps for ADHD have the ability to help improve their memory and inhibition impairments, thus targeting the impulsivity in ADHD. It was noted that the apps themselves could be a distraction to other important tasks (n=1):

> *...if they're rewarded for each activity...they're going to engage...cos they're gonna get their dopamine hit each time.* [HCP4]

Age Appropriateness

Young people (n=4) and clinicians (n=5) believed that it is important for the apps to be age appropriate for the user. Some participants said that some apps were a bit "babyish," and some noted that one of the apps was a bit too old for a child because of the language used. One participant noted that one of the apps was describing a scene that was "hard to imagine" (YP1). The participants noted:

> *A bit childish, does it say for 13 [years old]?* [YP2]
>
> *...I'd rather have it when I'm older.* [YP4]
>
> *...it doesn't feel as if it's aimed at kids.* [HCP3]
>
> *I think it's very immature, I dunno what the age range is but it feels like the cartoons I used to watch when I was little *laughs*.* [HCP5]

App Interaction

Young people stated that apps must be interactive (n=3). They were less impressed and, at times, frustrated with apps that involved no interaction but simply listening or watching (n=3). They wanted to be involved and interact with the app itself, as noted below:

> *...how I didn't do anything! This app is fake! Its fake! [the app said he had achieved the next level and he hadn't done anything].* [YP3]

> *I didn't like it seriously, it's mind controlling! [as above, the app said she had achieved the next level and he hadn't done anything].* [YP5]
>
> *I would like change it a bit so you can like actually do something cos and the moment on the writing you just you don't...do anything...* [YP4]

One clinician also noted that the apps should incorporate learning and make it fun, have a catchy title and app logo, and a simple description. One person highlighted that ADHD medication could affect a child's ability to engage with an app:

> *Some children who we see with ADHD will be medicated and some won't and that will affect their levels of concentration...* [HCP2]

The following three themes were considered to be important by clinicians but not mirrored by the young people:

Monitoring

Clinicians suggested that apps could be used to monitor diet (n=2), a "mood component" (C1; n=2), and ADHD symptoms within this population. Some believed they would be useful for keeping a diary (n=3) of these, and one person believed apps would be useful to act as ADHD medication reminders:

> *A child and parent app then the parent can then monitor [ADHD symptoms].* [HCP1]
>
> *I think something about a diary in terms of perhaps providing prompts or things that they've got going on that day...* [HCP2]
>
> *I think it's really good to encourage kids with ADHD to monitor the way they feel.* [HCP3]

App Effect on Relationships

One clinician was mindful that the use of apps could prompt "disharmony" in the home, causing conflict in terms of sharing with siblings (C2). They were also mindful that a number of the apps encourage the user to access them daily. This clinician believed that this could encourage a parent to nag their child to use the app, causing tension in the home as well:

> *...it kind of might at some point cause tension and then, you know disharmony at home and arguments and things like that, prompting them to try and get them to do.* [HCP2]

Others believed that as one common difficulty in ADHD is maintaining relationships with friends and family, apps could help facilitate these relationships rather than be harmful to them (n=2):

> *...err I think relationships with peers would be useful...* [HCP2]
>
> *I think peer relationships would definitely be one thing...if you're looking at relationships you could also mention family relationships as well.* [HCP1]

The Impact of Common Comorbid Conditions

Additionally, they highlighted that ADHD is a highly comorbid condition and the apps don't always account for this (n=2). One clinician shared that children with ADHD may have learning difficulties and be behind academically and that some of the apps used appeared challenging for the target age even without the young people having additional needs. One participant stated that autism spectrum disorders are a common ADHD comorbidity and that a symptom of ASD can be the inability to cope with multiple stimuli at once. For example, some of the apps were very visually "busy" with lots of background noise. This clinician described this as "sensory overload" (HCP4); something these users may not always be able to cope with:

> *...a lot of children as you know with ADHD have got autism as well and so sometimes things that are completely too noisy, too bright, unless they pick them, can be a bit of a problem.* [HCP4]

Discussion

Principal Findings

This study identified 10 apps that stated they were aimed at children and/or young people with ADHD. Five clinicians working with this population and 5 young people with ADHD were interviewed. Interviews involved sharing the 10 identified apps with participants and asking what they think would make an app suitable for children and young people with ADHD. Young people stated that technology unreliability can be a barrier when trying to engage them with the technology; young people and clinicians believe that the young people want to be able to relate to an app, clinicians wanted apps to target the ADHD symptoms of young people, young people and clinicians wanted apps to provide rewards, and young people wanted to be able to interact with an app. Additionally, 2 clinicians felt apps would be useful to monitor ADHD symptoms, diet, and mood and to improve family and peer relationships.

Conclusions

ADHD is a chronic condition with around 60% going on to have some symptoms in adult life [2]. Treatment with medication for ADHD falls significantly during the second decade of life [54]—a time when many young people with ADHD struggle with school, family, and peer relationships and risks such as drug misuse. As with all chronic conditions, empowering patients to manage their own symptoms is crucial. Multimodal treatment is recommended for ADHD management [20], but in most countries, resources for ongoing nonpharmacological support are often thin.

Individuals with ADHD, and especially those with comorbid ASD, often enjoy technology and indeed are skilled with its use. Many apps are advertised for individuals with ADHD and their families [36,37,40,42-44], but little is known about their use and value. In this small study, we aimed to collect the views of the target group, children and young people with ADHD, and of clinicians who specialize in management of ADHD on what qualities they would like to see in an app and for ideas on when apps could be useful. The young people and clinicians whose views were sought in this study did not find that the apps reviewed fully met their expectations. The highest match was 5 out of 8; two apps scored 0 and the mean was 2.3 out of 8. Young people wanted to enjoy the technology and to use it for some specific tasks. The clinicians were keen to explore apps as a way to engage young people, motivate them to manage

their difficulties, and target specific needs for particular young people.

The perspectives of both the clinicians and the young people are needed to build up a full picture. It is important to have the views of clinicians as to whether the apps promote positive behaviors and target specific needs. The views of the young people give guidance as to whether they would actually use the app and whether they found it enjoyable and/or helpful. Individuals with ADHD show most reliable responses to frequent positive reward, so this seems to be a prerequisite of a successful app for the young people themselves.

Both clinicians and their patients wanted apps to be technically reliable, to relate in some way to the user, to be age appropriate, and to be interactive. Both were also keen to exploit technology to help them, including, for example, an app to address ADHD symptoms in some way. App developers may advertise their apps to offer what clinicians and young people want, but as yet, there does not seem to be a real evidence base to help families and clinicians decide whether or not an app is likely to work for them. Claims that apps can improve ADHD symptoms need further exploring. Clinical networks and children, young people, and families could work with app developers to trial existing apps and develop new ones.

Clinicians also saw the use of apps as a way of collecting data from their patients. For example, some young people do not gain weight on ADHD medications [55]. An app could perhaps be used to remind a young person to eat and to record what, if anything, they do eat. If a young person is suffering from low mood, perhaps an app could be used to record this and share mood scores with the clinician. It can be extremely difficult for a young person with ADHD to reliably report their dietary intake or their mood over time when they attend appointments [56].

The marketplace for app development and specifically app development in health care continues to expand. It is easy to see that conflict could exist between the desire to make money from an app by producing it in a less-than-ideal way and marketing it quickly and spending time and money researching the benefits of an app. Indeed, once an app has been purchased, the developer may move on to the next project, and families and patients who purchase cheap apps are unlikely to complain if they do not work well for them.

There is a need for guidelines and standards for app developers, including a requirement for transparent trials of the usefulness of the app. Clearly, there is however, a balance between making these so time consuming and expensive as to deter developers.

To investigate the factors influencing technology acceptance, a model named the technology acceptance model (TAM) provides a basis for attitude measures with two technology acceptance variables: perceived usefulness (PU) and perceived ease of use (PEU) [57]. PU refers to “the degree to which a person believes that using a particular system would enhance his or her job performance.” PEU is defined as “the degree to which a person believes that using a particular system would be free from effort.” Research has shown that TAM has been one of the most influential models in explaining user acceptance of information technology (IT), and it has gained wide attention in the IT literature because it includes the psychological interaction of a user with technology (unpublished data, 2009 [58]). According to TAM, if users perceive a technology as useful and easy to use, they develop positive attitudes toward the technology.

Future research is needed on the value of apps for children and young people with ADHD and their families and, in particular, any positive role for apps in the management of ADHD in this age group. A systematic review on how technology can be used to engage this population and how it can be used to help them would be a useful way forward. This could be the platform to begin exploring the use of apps further.

Acknowledgments

This research was funded and supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber. The views and opinions expressed are those of the authors and are not necessarily those of the NHS, the NIHR, or the Department of Health, United Kingdom.

This study was funded by the University of Sheffield’s Teaching Assistant Sheffield undergraduate Research Experience Scheme and the Engineering and Physical Sciences Research Council (EPSRC) Vacation Bursary Scheme.

The authors would like to thank the children, young people, their families, and the clinicians for taking part and sharing their expertise.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Information about apps in this research.

[PDF File (Adobe PDF File), 14KB - mhealth_v5i10e145_app1.pdf]

References

1. Biederman J. Attention-deficit/hyperactivity disorder: a selective overview. Biol Psychiatry 2005 Jun 01;57(11):1215-1220. [doi: 10.1016/j.biopsych.2004.10.020] [Medline: 15949990]
2. Faraone SV, Biederman J, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. Psychol Med 2006 Feb;36(2):159-165. [doi: 10.1017/S003329170500471X] [Medline: 16420712]
3. Harpin VA. The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. Arch Dis Child 2005 Feb;90(Suppl 1):i2-i7 [FREE Full text] [doi: 10.1136/adc.2004.059006] [Medline: 15665153]
4. Biederman J, Newcorn J, Sprich S. Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. Am J Psychiatry 1991 May;148(5):564-577. [doi: 10.1176/ajp.148.5.564] [Medline: 2018156]
5. Kadesjö B, Gillberg C. The comorbidity of ADHD in the general population of Swedish school-age children. J Child Psychol Psychiatry 2001 May;42(4):487-492. [Medline: 11383964]
6. Yoshimasu K, Barbaresi WJ, Colligan RC, Voigt RG, Killian JM, Weaver AL, et al. Childhood ADHD is strongly associated with a broad range of psychiatric disorders during adolescence: a population-based birth cohort study. J Child Psychol Psychiatry 2012 Oct;53(10):1036-1043 [FREE Full text] [doi: 10.1111/j.1469-7610.2012.02567.x] [Medline: 22647074]
7. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry 2007 Jun;164(6):942-948. [doi: 10.1176/ajp.2007.164.6.942] [Medline: 17541055]
8. Antshel KM, Faraone SV, Maglione K, Doyle A, Fried R, Seidman L, et al. Temporal stability of ADHD in the high-IQ population: results from the MGH longitudinal family studies of ADHD. J Am Acad Child Adolesc Psychiatry 2008 Jul;47(7):817-825. [doi: 10.1097/CHI.0b013e318172eecf] [Medline: 18520956]
9. Barkley RA, Fischer M, Edelbrock CS, Smallish L. The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. J Am Acad Child Adolesc Psychiatry 1990 Jul;29(4):546-557. [doi: 10.1097/00004583-199007000-00007] [Medline: 2387789]
10. Biederman J, Faraone S, Milberger S, Curtis S, Chen L, Marrs A, et al. Predictors of persistence and remission of ADHD into adolescence: results from a four-year prospective follow-up study. J Am Acad Child Adolesc Psychiatry 1996 Mar;35(3):343-351 [FREE Full text] [doi: 10.1097/00004583-199603000-00016] [Medline: 8714323]
11. Biederman J, Monuteaux MC, Mick E, Spencer T, Wilens TE, Silva JM, et al. Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. Psychol med 2006;36(2):167-179 [FREE Full text] [Medline: 16420713]
12. Biederman J, Petty CR, Evans M, Small J, Faraone SV. How persistent is ADHD? a controlled 10-year follow-up study of boys with ADHD. Psychiatry Res 2010 May 30;177(3):299-304 [FREE Full text] [doi: 10.1016/j.psychres.2009.12.010] [Medline: 20452063]
13. Gittelman R, Mannuzza S, Shenker R, Bonagura N. Hyperactive boys almost grown up. I. Psychiatric status. Arch Gen Psychiatry 1985 Oct;42(10):937-947. [Medline: 4037987]
14. Langley K, Fowler T, Ford T, Thapar AK, van den Bree M, Harold G, et al. Adolescent clinical outcomes for young people with attention-deficit hyperactivity disorder. Br J Psychiatry 2010 Mar 01;196(3):235-240. [doi: 10.1192/bjp.bp.109.066274] [Medline: 20194547]
15. McCormick LH. Adult outcome of child and adolescent attention deficit hyperactivity disorder in a primary care setting. South Med J 2004 Sep;97(9):823-826. [doi: 10.1097/01.SMJ.0000129931.63727.10] [Medline: 15455963]
16. Mick E, Byrne D, Fried R, Monuteaux M, Faraone SV, Biederman J. Predictors of ADHD persistence in girls at 5-year follow-up. J Atten Disord 2011 Apr;15(3):183-192 [FREE Full text] [doi: 10.1177/1087054710362217] [Medline: 20332414]
17. Barbaresi WJ, Colligan RC, Weaver AL, Voigt RG, Killian JM, Katusic SK. Mortality, ADHD, and psychosocial adversity in adults with childhood ADHD: a prospective study. Pediatrics 2013 Apr;131(4):637-644 [FREE Full text] [doi: 10.1542/peds.2012-2354] [Medline: 23460687]
18. Mannuzza S, Klein RG, Bessler A, Malloy P, LaPadula M. Adult psychiatric status of hyperactive boys grown up. Am J Psychiatry 1998 Apr;155(4):493-498. [doi: 10.1176/ajp.155.4.493] [Medline: 9545994]
19. Taylor E, Chadwick O, Heptinstall E, Danckaerts M. Hyperactivity and conduct problems as risk factors for adolescent development. J Am Acad Child Adolesc Psychiatry 1996 Sep;35(9):1213-1226. [doi: 10.1097/00004583-199609000-00019] [Medline: 8824065]
20. NICE. London, UK; 2008. Attention deficit hyperactivity disorder: diagnosis and management URL: https://www.nice.org.uk/guidance/CG72 [accessed 2017-09-11] [WebCite Cache ID 6tOXKy43k]
21. Parker J, Wales G, Chalhoub N, Harpin V. The long-term outcomes of interventions for the management of attention-deficit hyperactivity disorder in children and adolescents: a systematic review of randomized controlled trials. Psychol Res Behav Manag 2013 Sep 17;6:87-99 [FREE Full text] [doi: 10.2147/PRBM.S49114] [Medline: 24082796]
22. Visser SN, Danielson ML, Bitsko RH, Holbrook JR, Kogan MD, Ghandour RM, et al. Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003-2011. J Am Acad Child Adolesc Psychiatry 2014 Jan;53(1):34-46.e2 [FREE Full text] [doi: 10.1016/j.jaac.2013.09.001] [Medline: 24342384]
23. Taylor K. Deloitte: London. London: Deloitte; 2015. Connected heath, how digital technology is transforming health and social care URL: https://www2.deloitte.com/uk/en/pages/life-sciences-and-healthcare/articles/connected-health.html [accessed 2017-01-12] [WebCite Cache ID 6nSntAqrp]

24. Xu C, Reid R, Steckelberg A. Technology applications for children with ADHD: assessing the empirical support. Educ Treat Children 2002 May;25(2):224-248.
25. Drigas AS, Ioannidou RE, Kokkalia G, Lytras MD. ICTs, mobile learning and social media to enhance learning for attention difficulties. Journal Univers Comput Sci 2014 Jun 20;20(10):1499-1510. [doi: 10.3217/jucs-020-10-1499]
26. Shrieber B, Seifert T. College students with learning disabilities and/or ADHD use of a handheld computer compared to conventional planners. In: The Open University of Israel. 2009 Presented at: Proceedings of the Chais conference on instructional technologies research 2009; 2009; Israel.
27. Gulchak DJ. Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. Educ and Treat Children 2008;31(4):567-581.
28. Hecker L, Burns L, Katz L, Elkind J, Elkind K. Benefits of assistive reading software for students with attention disorders. Ann Dyslexia 2002;52(1):243-272.
29. Ota KR, DuPaul GJ. Task engagement and mathematics performance in children with attention-deficit hyperactivity disorder: effects of supplemental computer instruction. Sch Psychol Q 2002;17(3):242-257. [doi: 10.1521/scpq.17.3.242.20881]
30. Bul KC, Franken IH, Van der Oord S, Kato PM, Danckaerts M, Vreeke LJ, et al. Development and user satisfaction of "Plan-It Commander, " a serious game for children with ADHD. Games Health J 2015 Dec;4(6):502-512. [doi: 10.1089/g4h.2015.0021] [Medline: 26325247]
31. Edebol H, Helldin L, Norlander T. Measuring adult attention deficit hyperactivity disorder using the quantified behavior test plus. Psych J 2013 Apr;2(1):48-62 [FREE Full text] [doi: 10.1002/pchj.17] [Medline: 24294490]
32. Hall CL, Walker GM, Valentine AZ, Guo B, Kaylor-Hughes C, James M, et al. Protocol investigating the clinical utility of an objective measure of activity and attention (QbTest) on diagnostic and treatment decision-making in children and young people with ADHD-'Assessing QbTest Utility in ADHD' (AQUA): a randomised controlled trial. BMJ Open 2014 Dec 01;4(12):e006838 [FREE Full text] [doi: 10.1136/bmjopen-2014-006838] [Medline: 25448628]
33. Gringras P, Santosh P, Baird G. Development of an internet-based real-time system for monitoring pharmacological interventions in children with neurodevelopmental and neuropsychiatric disorders. Child Care Health Dev 2006 Sep;32(5):591-600. [doi: 10.1111/j.1365-2214.2006.00653.x] [Medline: 16919139]
34. Livingstone S, Haddon L, Görzig A, Ólafsson K. The London School of Economics and Political Science. London: EU Kids Online; 2011. Risks and safety for children on the internet: the UK report: full findings from the EU Kids Online survey of UK 9-16 year olds and their parents URL: http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20II%20(2009-11)/National%20reports/UKReport.pdf [accessed 2017-01-19] [WebCite Cache ID 6ndTYe6Q8]
35. Young Z, Craven MP, Groom M, Crowe JA. Snappy App: a mobile continuous performance test with physical activity measurement for assessing attention deficit hyperactivity disorder. 2014 Presented at: International Conference on Human-Computer Interaction; 2014; Switzerland.
36. Chen H, Yang HI, Hooks H, Lee J, Satterfield D, Wong J, et al. Medbuddy: a mobile medicinal management system for children with ADD/ADHD. 2012 Presented at: International Conference on Smart Homes and Health Telematics; 2012; Italy.
37. Ruiz-Manrique G, Tajima-Pozo K, Montañes-Rada F. Case report: "ADHD Trainer": the mobile application that enhances cognitive skills in ADHD patients. F1000Res 2014;3:283 [FREE Full text] [doi: 10.12688/f1000research.5689.3] [Medline: 26962432]
38. Sack E. Viurrspace. 2016. On-Task: increasing attentiveness in students with ADHD URL: https://viurrspace.ca/bitstream/handle/10170/884/sack_erich.pdf?sequence=1 [accessed 2017-09-18] [WebCite Cache ID 6tZB6fApx]
39. Molinero AA, Hernández FJ, Zorrilla AM, Zapirain BG. Technological solution for improving time management skills using an android application for children with ADD. 2012 Presented at: International Workshop on Ambient Assisted Living; 2012; Spain.
40. Pinna JE. Proquest. 2015. Development and preliminary testing of tablet application to increase reading motivation and summarization for adolescent students with ADHD URL: https://search.proquest.com/openview/1a7088139547180a193411af15f4a694/1.pdf?pq-origsite=gscholar&cbl=18750&diss=y
41. Schuck S, Emmerson N, Ziv H, Collins P, Arastoo S, Warschauer M, et al. Designing an iPad app to monitor and improve classroom behavior for children with ADHD: iSelfControl feasibility and pilot studies. PLoS One 2016;11(10):e0164229 [FREE Full text] [doi: 10.1371/journal.pone.0164229] [Medline: 27741257]
42. Spachos D, Chifari A, Chiazzese G, Merlo G, Doherty G, Bamidis PD. WHAAM: a mobile application for ubiquitous monitoring of ADHD behaviors. 2015 Presented at: International Conference of Interactive Mobile Communication Technologies and Learning (IMCL); 2014; Greece. [doi: 10.1109/IMCTL.2014.7011153]
43. Zuckeman O, Gal-Oz A, Tamir N, Kopelman-Rubin D. Initial validation of an assistive technology to enhance executive functioning among children with ADHD. 2015 Presented at: 14th International Conference on Interaction Design and Children; 2015; Boston.
44. Bamford C. Rowan University. 2016. The effects of self-monitoring using an iPad to increase on-task behaviors of high school students with attention deficit/hyperactivity disorder (ADHD) URL: http://rdw.rowan.edu/etd/1581/ [accessed 2017-09-14] [WebCite Cache ID 6tTCXTWpq]

45. Craven MP. Royal College of Psychiatrists. Cardiff: MindTech Healthcare Technology Cooperative; 2014. Technologies for remote therapy and management? multiple criteria, multiple stakeholders URL: http://www.rcpsych.ac.uk/pdf/Craven%20Michael%20P%20CAP14.pdf [accessed 2017-01-12] [WebCite Cache ID 6nSoVBIIS]
46. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
47. Astell AJ, Joddrell P, Groenewoud H, de Lange J, Goumans M, Cordia A, et al. Does familiarity affect the enjoyment of touchscreen games for people with dementia? Int J Med Inform 2016 Jul;91:e1-e8. [doi: 10.1016/j.ijmedinf.2016.02.001] [Medline: 26897552]
48. Joddrell P, Hernandez A, Astell AJ. Identifying existing, accessible touchscreen games for people living with dementia. 2016 Presented at: International Conference on Computers Helping People with Special Needs; 2016; Austria.
49. Statista. 2017. Number of apps available in leading app stores as of March 2017 URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2017-07-20] [WebCite Cache ID 6s5t9U605]
50. Cailean I. Trademob. 2015. Extensive guide to app store optimization (ASO) in 2015? part 1: apple app store URL: http://www.trademob.com/app-store-optimization-guide-apple/ [accessed 2017-01-19] [WebCite Cache ID 6ndTncE0B]
51. Cailean I. Trademob. 2015. Extensive guide to app store optimization (ASO) in 2015? part 2: google play store URL: http://www.trademob.com/app-store-optimization-guide-google/ [accessed 2017-01-19] [WebCite Cache ID 6ndTrjaCK]
52. Guest G, Bunce A, Johnson L. How many interviews are enough? an experiment with data saturation and variability. Field methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
53. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000;320(7227):114-116 [FREE Full text] [Medline: 10625273]
54. McCarthy S, Wilton L, Murray ML, Hodgkins P, Asherson P, Wong IC. The epidemiology of pharmacologically treated attention deficit hyperactivity disorder (ADHD) in children, adolescents and adults in UK primary care. BMC Pediatr 2012 Jun 19;12:78 [FREE Full text] [doi: 10.1186/1471-2431-12-78] [Medline: 22712630]
55. Charach A, Figueroa M, Chen S, Ickowicz A, Schachar R. Stimulant treatment over five years: effects on growth. J Am Acad Child Adolesc Psychiatry 2006;45(4):415-421 [FREE Full text]
56. Barkley RA, Fischer M, Smallish L, Fletcher K. The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. J Abnorm Psychol 2002 May;111(2):279-289. [Medline: 12003449]
57. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
58. Isık O. Turkish EFL teachers' attitudes towards ICT integration in language classrooms. WJEIS 2009;5:1-11 [FREE Full text]

Abbreviations

ADD: attention deficit disorder
ADHD: attention-deficit/hyperactivity disorder
ASO: app store optimization
NHS: National Health Service
NIHR: National Institute for Health Research
PEU: perceived ease of use
PU: perceived usefulness
QbTest: quantified behavior test
TAM: technology acceptance model

Edited by G Eysenbach; submitted 25.01.17; peer-reviewed by K Boydell, J Behrens; comments to author 12.07.17; revised version received 20.07.17; accepted 06.08.17; published 04.10.17.

Please cite as:
Powell L, Parker J, Robertson N, Harpin V
Attention Deficit Hyperactivity Disorder: Is There an App for That? Suitability Assessment of Apps for Children and Young People With ADHD
JMIR Mhealth Uhealth 2017;5(10):e145
URL: http://mhealth.jmir.org/2017/10/e145/
doi:10.2196/mhealth.7371
PMID:28978497

Original Paper

Mobile Apps for Suicide Prevention: Review of Virtual Stores and Literature

Isabel de la Torre[1], BE, MS, PhD; Gema Castillo[1], MSc; Jon Arambarri[2], BE, MS, PhD; Miguel López-Coronado[1], BE, MS, PhD; Manuel A Franco[3], BE, MS, PhD

[1]Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Valladolid, Spain

[2]VirtualWare Labs Foundation, Bilbao, Spain

[3]Psychiatry Service, Hospital of Zamora, Zamora, Spain

Corresponding Author:
Isabel de la Torre, BE, MS, PhD
Department of Signal Theory and Communications, and Telematics Engineering
University of Valladolid
Paseo de Belén 15
Valladolid, 47011
Spain
Phone: 34 983423000 ext 3703
Email: isator@tel.uva.es

Abstract

Background: The best manner to prevent suicide is to recognize suicidal signs and signals, and know how to respond to them.

Objective: We aim to study the existing mobile apps for suicide prevention in the literature and the most commonly used virtual stores.

Methods: Two reviews were carried out. The first was done by searching the most commonly used commercial app stores, which are iTunes and Google Play. The second was a review of mobile health (mHealth) apps in published articles within the last 10 years in the following 7 scientific databases: Science Direct, Medline, PsycINFO, Embase, The Cochrane Library, IEEE Xplore, and Google Scholar.

Results: A total of 124 apps related to suicide were found in the cited virtual stores but only 20 apps were specifically designed for suicide prevention. All apps were free and most were designed for Android. Furthermore, 6 relevant papers were found in the indicated scientific databases; in these studies, some real experiences with physicians, caregivers, and families were described. The importance of these people in suicide prevention was indicated.

Conclusions: The number of apps regarding suicide prevention is small, and there was little information available from literature searches, indicating that technology-based suicide prevention remains understudied. Many of the apps provided no interactive features. It is important to verify the accuracy of the results of different apps that are available on iOS and Android. The confidence generated by these apps can benefit end users, either by improving their health monitoring or simply to verify their body condition.

(JMIR Mhealth Uhealth 2017;5(10):e130) doi:10.2196/mhealth.8036

KEYWORDS

app; literature; prevention; suicide; virtual stores

Introduction

According to the World Health Organization, approximately one million people commit suicide each year, which is the second leading cause of death among people from 15 to 29 years of age [1]. There are also indications that for every adult who committed suicide, possibly more than 20 others attempted suicide [1]. Worldwide, suicides account for 50% of all violent deaths recorded among men and 71% among women [2]. In Spain, suicide causes twice as many deaths as traffic accidents and 70 times more deaths than gender violence [1].

It is important to highlight the fact that suicides are preventable; prevention efforts would require a comprehensive multi-sectorial prevention response in all countries. Based on the societal impacts that suicide has in many countries, it is very important that health services incorporate suicide prevention as a central component. In addition, alcohol consumption and mental disorders contribute to many suicides worldwide [2-4].

Early identification and effective management are key for people at risk for suicide, so they can receive the care they need [5]. Communities play a key role in suicide prevention: they can provide social support to the most vulnerable individuals, deal with their follow-up, fight against social stigmatization, and support those who have lost close relatives to suicide [6,7].

People feel anxiety and some situations can cause them to make decisions to end their life. Information and communications technologies (ICTs) can provide innovations to support the work of health professionals to address this issue. Mobile apps can also provide a valuable resource for communities and individuals by addressing social issues [3,4]. The aim of this article is to conduct a systematic review of existing mobile apps in Google and Apple online stores, and in the literature for suicide prevention. This research will help experts in the field and app developers gain insight into what mobile apps exist, and features that they can continue to research and develop.

Mobile health (mHealth) is a part of electronic health and refers to computers, smartphones, mobile health apps, and patient monitoring services and apps used for the purpose of positively increasing access to health information. For many health issues, especially those related to mental health prevention, remote patient monitoring is a good start and mobile apps on smartphones are an available means that can save costs and benefit hard-to-reach populations. However, it is important to verify the accuracy of the results of different apps that are available on the virtual stores, such as iOS [8] and Android [9]. The confidence generated by these apps can benefit end users, either by improving their health monitoring or simply to verify their body condition.

The goal of this paper is to present the state of the art of mobile apps designed to prevent suicide, in English and Spanish languages. This review of the scientific literature is presented to show scientific studies that demonstrate how such apps have been developed, and the results that have been obtained, when using mobile apps. This review allows us to promote the discussion of the results found, and to communicate the current state of these technologies to other members of the scientific community.

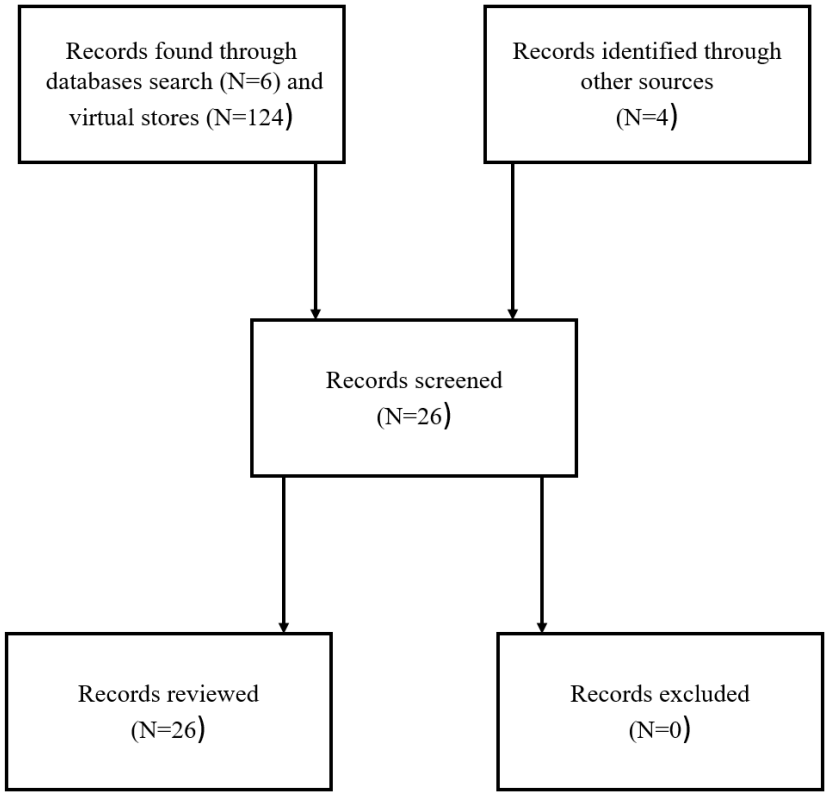
Methods

Search Strategy for the Reviews

Two reviews have been completed; the first was carried out by searching the most commonly used commercial apps stores, which are Google Play [9] and iTunes [8]. The second review was a search of published articles from the following scientific databases: Science Direct, Medline, PsycINFO, Embase, The Cochrane Library, IEEE Xplore, and Google Scholar.

The search terms used for virtual stores were “*prevention*” and “*suicide*”. Only apps that focused on suicide prevention were studied; music, games, and other apps were dismissed. The literature review was developed on the indicated scientific databases. The combination of terms in titles/abstracts used in the literature revision was the following: “*prevention*” AND “*suicide*” AND “*app/mhealth*”. The results were limited to the last 10 years, from 2007 forward. Only articles published in English were studied. Figure 1 presents a flowchart with the steps that were followed in the review.

Figure 1. Flow chart of the steps in the review.



Results

Virtual Store Apps

A total of 20 relevant apps were found in the Google Play and Apple App Stores, although we note that some apps were available in both stores. A classification of these apps is shown in Multimedia Appendix 1. The characteristics chosen were: name, operating system, language, brief description, and price. The number of downloads has not been taken into account, because this value can change continuously. Only one app was found in Spanish (Prevensuic), which was designed by the Spanish Mental Health Foundation for the prevention of mental disorders and suicide. This app was developed for Android and iOS.

Apps in the Literature

A total of 6 relevant papers were found in the online databases that were searched. Shand et al [10] evaluated the effectiveness of a self-help app for suicidal thoughts amongst young Indigenous people (with a total of 150 participants). This research was the first to evaluate the effectiveness of a self-help app for suicidal thoughts amongst young Indigenous people.

Aguirre et al [11] described mobile apps for one specific field of Human Service Organizations related to suicide prevention, and other topics in mental health and behavioral issues. This study analyzed 27 apps, and concluded that it was critical to carefully plan and develop the apps, and undertake ongoing evaluation of these apps [11]. Berrouiguet et al [12] proposed a mobile app for suicide prevention that was a connected tool used by the patient to describe his/her health status. The authors concluded that the app should be ergonomic, with data protection and question choices [12]. These facts should be considered in the future to increase acceptance by both patients and practitioners.

Larsen et al [13] found 123 apps referring to suicide, but not all were specifically related to suicide prevention. The authors warned that physicians should be wary in recommending apps [13]. McManama et al [14] evaluated a smartphone app intervention developed specifically for suicidal adolescents and their parents, named *Crisis Care*. Twenty adolescent-parent dyads participated in the pilot evaluation, and the results showed positive results for usability, acceptability, and utility of the app [14]. This app can be an adjunct to treatment for suicidal adolescents and their parents [14].

Kennard et al [15] showed the results from phase one of a treatment-development study for suicide prevention. This study used qualitative methods to collect information from families, teens, and physicians and evaluated acceptability and the feasibility to provide the intervention to patients and parents. All participants were constructive about the utility of technology in safety planning with the mHealth app, but limitations included privacy and confidentiality issues [15].

Discussion

Suicide prevention is a health issue, and it is important that apps in this field are supported by professionals who make these tools as a way to care for their patients, and to prevent this situation that affects many countries.

Few commercial mobile apps aimed at suicide prevention exist, compared to other pathologies such as cardiology [16]. As far as existing literature is concerned, even fewer results have been found. Luxton et al [17] examined innovate apps related to suicide prevention. Some of these apps included gaming, text analysis, and virtual worlds, and the authors discussed the limitations and advantages of these technologies [17].

Marasinghe et al [18] conducted a randomized controlled trial examining whether a *Brief Mobile Treatment* intervention could improve outcomes related to typical care among people with suicidal behavior. Sixty-eight participants were recruited from a Sri Lankan hospital; people who received the intervention were found to accomplish important improvements in reducing suicidal ideation [18]. Larsen et al [19] showed an overview of different technological developments for suicide prevention. Some of these developments included automatic detection of suicide cases from social media and crisis detection from acoustic variability, among others [19]. The authors presented the mHealth app for Indigenous populations that was detailed in Shand et al [10]. Barriers to help-seeking included shame, feared loss of autonomy, and negative attitudes towards health care providers [10]. The use of mobile devices and apps continues to rise amongst young people, thus presenting opportunities to utilize these aids in overcoming help-seeking barriers.

Based on our experience in clinical and real cases, we conclude that mHealth apps are an essential tool that can help us to prevent suicide, considering the family, caregivers, and health professionals that are part of helping to save a life through the use of ICTs. Therefore, we consider it fundamental that clinical support provides validity to the apps offered on the Internet. The bibliographic review allowed us to examine the apps based on scientific studies from real cases that generated proven results.

As future work, we will develop and evaluate a suicide prevention app by looking at the shortcomings detected in the apps that were analyzed in this work. Saving even a single life via a mobile app is a breakthrough.

Acknowledgments

This research was partially supported by the European Commission and the Ministry of Industry, Energy, and Tourism under the project AAL-20125036 named *WetakeCare: ICT-based Solution for (Self-) Management of Daily Living*.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Suicide prevention apps in virtual stores.

[PDF File (Adobe PDF File), 35KB - mhealth_v5i10e130_app1.pdf]

References

1. World Health Organization. Suicide data. 2017. URL: http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ [accessed 2017-08-17] [WebCite Cache ID 6smNBJWkg]
2. Mathias C, Michael FR, Sheftall A, Hill-Kapturczak N, Crum P, Dougherty D. What's the harm in asking about suicidal ideation? Suicide Life Threat Behav 2012 Jun;42(3):341-351 [FREE Full text] [doi: 10.1111/j.1943-278X.2012.0095.x] [Medline: 22548324]
3. Siegel A. Suicide prevention by smartphone. Am J Med Internet 2016 Aug;129:E145 [FREE Full text] [doi: 10.1016/j.amjmed.2016.01.034] [Medline: 27453384]
4. Lai MH, Maniam T, Chan LF, Ravindran AV. Caught in the web: a review of web-based suicide prevention. J Med Internet Res 2014;16(1):e30 [FREE Full text] [doi: 10.2196/jmir.2973] [Medline: 24472876]
5. Ghoncheh R, Gould MS, Twisk JW, Kerkhof AJ, Koot HM. Efficacy of adolescent suicide prevention e-learning modules for gatekeepers: a randomized controlled trial. JMIR Ment Health 2016 Jan 29;3(1):e8 [FREE Full text] [doi: 10.2196/mental.4614] [Medline: 26825006]
6. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing messaging to engage patients in an online suicide prevention intervention: survey results from patients with current suicidal ideation. J Med Internet Res 2014 Feb 07;16(2):e42 [FREE Full text] [doi: 10.2196/jmir.3173] [Medline: 24509475]
7. 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]
8. Apple Inc.. iTunes. 2017. URL: https://www.apple.com/itunes/ [accessed 2017-08-17] [WebCite Cache ID 6smNlFSwP]
9. Google. App in Android - Google Play. 2017. URL: https://play.google.com/store/search?q=suicide&c=apps [accessed 2018-08-28] [WebCite Cache ID 6t3yUIcjs]
10. Shand FL, Ridani R, Tighe J, Christensen H. The effectiveness of a suicide prevention app for indigenous Australian youths: study protocol for a randomized controlled trial. Trials 2013;14:396 [FREE Full text] [doi: 10.1186/1745-6215-14-396] [Medline: 24257410]
11. Aguirre RT, McCoy MK, Roan M. Development guidelines from a study of suicide prevention mobile applications (apps). J Technol Hum Serv 2013;31(3):293. [doi: 10.1080/15228835.2013.814750]
12. Berrouiguet S, Billot R, Lenca P, Tanguy P, Baca-Garcia E, Simonnet M. Toward e-health applications for suicide prevention. 2016 Jun 27 Presented at: 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); June 27-29, 2016; Washington, DC, USA p. 346. [doi: 10.1109/CHASE.2016.37]
13. Larsen M, Nicholas J, Christensen H, Beautrais A, Currier D, Haas A. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016;11(4):e0152285. [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
14. McManama OK, LeCloux M, Ross A, Gironda C, Wharff EA. A pilot study of the acceptability and usability of a smartphone application intervention for suicidal adolescents and their parents. Arch Suicide Res 2017;21(2):254-264. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
15. 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]
16. 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]
17. Luxton D, June J, Kinn J. Telemedicine and e-health. 2011. Brief communication. Technology-based suicide prevention: current applications and future directions URL: https://www.iasp.info/pdf/50-54.pdf [accessed 2017-08-28] [WebCite Cache ID 6t3yyz3BB]
18. 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]
19. Larsen ME, Cummins N, Boonstra TW, O'Dea B, Tighe J, Nicholas J, et al. The use of technology in suicide prevention. Conf Proc IEEE Eng Med Biol Soc 2015;2015:7316-7319. [doi: 10.1109/EMBC.2015.7320081] [Medline: 26737981]

Abbreviations

ICT: information and communications technology
mHealth: mobile health

Edited by G Eysenbach; submitted 12.05.17; peer-reviewed by J Herreros, R Dawson; comments to author 12.07.17; revised version received 19.07.17; accepted 04.08.17; published 10.10.17.

Please cite as:
de la Torre I, Castillo G, Arambarri J, López-Coronado M, Franco MA
Mobile Apps for Suicide Prevention: Review of Virtual Stores and Literature
JMIR Mhealth Uhealth 2017;5(10):e130
URL: https://mhealth.jmir.org/2017/10/e130/
doi:10.2196/mhealth.8036
PMID:29017992

Original Paper

ADHD: Is There an App for That? A Suitability Assessment of Apps for the Parents of Children and Young People With ADHD

Lauren Powell[1], BSc (Hons), MSc; Jack Parker[1], BSc (Hons), PhD; Valerie Harpin[2], MBBChir, MD, FRCP, FRCPCH

[1]School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom

[2]Sheffield Children's Hospital NHS Foundation Trust, Ryegate Children's Centre, Sheffield, United Kingdom

Corresponding Author:
Lauren Powell, BSc (Hons), MSc
School of Health and Related Research
University of Sheffield
Regent Court
30 Regent Street
Sheffield, S1 4DA
United Kingdom
Phone: 44 114 222 8275
Email: l.a.powell@sheffield.ac.uk

Abstract

Background: Attention-deficit hyperactivity disorder (ADHD) is a highly comorbid disorder that can impact significantly on the individual and their family. ADHD is managed via pharmacological and nonpharmacological interventions. Parents also gain support from parent support groups, which may include chat rooms, as well as face-to-face meetings. With the growth of technology use over recent years, parents have access to more resources that ever before. A number of mobile apps have been developed to help parents manage ADHD in their children and young people. Unfortunately many of these apps are not evidence-based, and little is known of their suitability for the parents or whether they are helpful in ADHD management.

Objective: The aim of this study was to explore the (1) parents' views of the suitability of the top ten listed apps for parents of children and young people with ADHD and (2) the views of clinicians that work with them on the suitability and value of the apps.

Methods: The top 10 listed apps specifically targeted toward the parents of children and young people with ADHD were identified via the Google Play (n=5) and iTunes store (n=5). Interviews were then undertaken with 7 parents of children or young people with ADHD and 6 clinicians who specialize in working with this population to explore their opinions of the 10 apps identified and what they believe the key components are for apps to be suitable and valuable for this population.

Results: Four themes emerged from clinician and parent interviews: (1) the importance of relating to the app, (2) apps that address ADHD-related difficulties, (3) how the apps can affect family relationships, and (4) apps as an educational tool. Two additional themes emerged from the clinician interviews alone: monitoring ADHD symptoms and that apps should be practical. Parents also identified an additional theme: the importance of the technology. Overall, the characteristics of the current top 10 listed apps did not appear to match well to the views of our sample.

Conclusions: Findings suggest that these apps may not fully meet the complex needs of this parent population. Further research is required to explore the value of apps with this population and how they can be tailored to their very specific needs.

(JMIR Mhealth Uhealth 2017;5(10):e149) doi:10.2196/mhealth.7941

KEYWORDS

attention deficit disorder with hyperactivity; mobile applications; technology

Introduction

ADHD in Children and Young People

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder, characterized by three core symptoms: hyperactivity, impulsivity, and inattention, which has a profound impact on the individual and their family [1-3]. ADHD is highly comorbid [4-6] and has a prevalence of 3% to 5% of school-aged children worldwide [7]. Furthermore, 80% to 85% of these children continue to be impaired by their ADHD

symptoms as adolescents [1,8-10] and 60% as adults [2]. The presence of ADHD also increases the risk of premature death [11-13]. When ADHD persists into adulthood, affected individuals are more likely to engage in criminality and substance abuse [14-16]. ADHD also places a large economic burden on society. In the United Kingdom, in 2010, it was estimated that ADHD in adolescents alone cost £670 million to the National Health Service (NHS), social care, and education resources [17]. The total annual cost of ADHD for children in the United States is estimated at US $38 to $72 billion [18]. ADHD in children and young people (YP) is also extremely challenging for parents to manage [3].

Globally, ADHD treatments include pharmacological and nonpharmacological interventions [19-21]. In less severe cases, nonpharmacological behavioral interventions such as psychoeducation programs, behavioral interventions, and cognitive behavioral therapy are used alone, and in more severe cases, it is recommended that both approaches are used in parallel [19].

When young people diagnosed with ADHD reach their adolescent years, they are less likely to take their ADHD medication reliably and are more likely to disengage from services [22]. This trend in young people is mirrored across other services such as diabetes [23] and mental health services [24-28]. This is problematic for a number of reasons including the negative effect ADHD symptoms can have on academic attainment [4,29], which can often result in the young people having to repeat a school year [30]. Adolescents are more likely to be expelled from school [31] and spend less time in education because of truancy than their peers [32]. All of these problems pose huge challenges for parents.

The Role of Parents in Child and Adolescent ADHD Management

Parents are under tremendous strain when caring for a child or YP with ADHD [3]. It is for this reason that the National Institute for Health and Care Excellence [19] guidelines recommend that all parents of children and young people diagnosed with ADHD are invited to attend parent management groups where they can build up a repertoire of strategies to help them manage their child's ADHD and help their child achieve in life and at school [19].

Emphasis has been placed on the importance of the supportive role of the parent in ADHD [3,33]. Young people with ADHD often underreport their ADHD-related difficulties compared with their parents [34]. Therefore it is important that they have the support of somebody such as a parent or guardian to assist with their difficulties, as the young people doesn't always appreciate how much difficulty they are having. Indeed, it is often the case that the parents require support for themselves so that they can provide optimal support for their child with ADHD. Available parenting interventions for the management of ADHD in primary school–aged children include an eight-session parenting behavioral program [35], the New Forest Parenting Programme [36], 123 Magic [37], and the Triple P Positive Parenting Program [38]. These programs involve strategies that cover aspects such as the understanding of ADHD and the challenging behaviors associated with it, specific behavioral strategies that address parent-child interaction, the use of time-out strategies to reduce problematic behavior, how to manage child behavior in public, and school and maintenance issues. Parent groups provide vital support for parents of children with ADHD and provide the parent with strategies that they can apply to managing their child's ADHD outside the group. However, many of them use paper-based resources which may not provide instant advice at a time where a parent needs it and management strategies will usually need reinforcement over time.

The Role of Technology in ADHD Management

The increasing use and accessibility of the Internet, mobile phones, and mobile devices offers an additional way to support parents [39]. Technology could be used for parents to engage their child in daily activities, monitor their symptoms, or to gain further advice on how to manage their child's or young person's ADHD. A recent qualitative study has demonstrated that parents, children, young people with ADHD, and health care professionals appreciate the potential of technology to help with the ongoing management and monitoring of a child's ADHD [40].

There have been a number of apps developed for parents for this purpose, but thus far, there is little evidence to underpin them. Some of these apps are designed for the parent and child to use jointly to enable the parent to monitor their child's ADHD symptoms, monitor medication compliance, give parenting advice for managing a child with ADHD, and also to help with daily routines such as getting dressed and going to bed. Evidence includes app development papers such as the WHAAM app that involves monitoring ADHD behavior in children and data sharing between parents, teachers, and health care professionals [41]. One study uses an app and a skin conductance sensor to measure parental stress by notifying the parent during stressful times to make them self-aware of their emotions. However, a number of false-positive alerts were apparent during this study [42]. Another study used a mobile app to help improve morning and bedtime routines. Parents reported decreased frustration levels in this study [43]. Nevertheless, these studies focus on the development of a single app and contain low sample sizes. This means we cannot apply these findings to a wider population.

Furthermore, app quality is not routinely monitored. Once the app has been bought, there is little feedback on usefulness. There is also minimal guidance available to demonstrate the reliability, validity [44], and suitability of available apps. For example, the British Standards Institution has developed a report outlining a code of practice recommended when developing a health app to help ensure it is fit for purpose [45]. This includes, and is not limited to, acknowledging and catering for the app's target audience. Unfortunately, many apps do not adhere to such guidelines. A review of the NHS Choices apps library was undertaken in England, and three components were focused upon: (1) data protection act compliance, (2) efficacy evidence, and (3) relevance to British people. Many apps that were identified in this review did not have an evidence base, and it was shown that privacy and data security arrangements were often unsuitable [46]. The thin evidence base for apps also

means that we know very little about what would make an app suitable for a parent of a child with ADHD.

This study proposes to identify apps aimed at the parents of children and young people with ADHD and to interview the parents and specialist clinicians who work with them to explore what could make an app suitable in terms of appearance, content, and functionality for the parents of children and young people with ADHD. This will provide those developing apps for this population with some key components to consider.

Methods

This research involved the identification of the top 10 listed apps aimed at the parents and guardians of children and young people diagnosed with ADHD. Subsequently, the parents tested the apps, and they were then interviewed to ascertain their views on the apps and to explore what they believed the key components are for apps to be helpful for them. Clinicians who work with children and young people with ADHD and their parents were also interviewed to explore their insights into how to make apps successful for this parent population.

Research Question

The research question was as follows: are the 10 top listed apps that are specifically designed and marketed for the parents of children and young people with ADHD suitable and what are the key components for apps to be suitable for this population?

Search and Identification of Mobile Apps

In July 2016, a search of mobile apps in the Apple iTunes store and the Android Google Play store was conducted. These databases were selected because they displayed systematically organized app rankings defined by algorithms unique to each app store, commonly known as app store optimization (ASO) [47]. For Apple, the primary factor is number of downloads; however there are also many other secondary factors such as keywords and visuals [48]. Similarly, the Android database is filtered according to multiple criteria, including the volume of ratings, value of ratings, and download growth [49]. Although this presents a potential bias as apps are selected according to the database's own ASO, it is to some extent unavoidable unless all of the search results are downloaded for testing [50], which was beyond the scope of this study. Apps for iPads and tablets were reviewed rather than phones, as apps are available on phone and tablet devices. For this study, apps available on tablets have been reviewed as they are easier to discuss with parents and clinicians.

The search term used was "ADHD." This is because other search terms such as "ADD," "children," "Young people," and "parents" did not provide different results to the "ADHD" searches. The term "ADHD" was searched in both iTunes and Google Play.

Preliminary screening was conducted based on app titles, full marketing description, and screenshots of the apps potentially relevant for inclusion. The first 5 listed apps that met the inclusion criteria from each app store were included, giving a total of 10 apps for study (Textboxes 1 and 2).

Duplicate apps were removed (Figure 1). This was applicable if there was more than one version of an app. It was decided that the app version to be included was the app that appeared first on the app store lists.

The remaining 10 apps were downloaded onto an Apple iPad Mini (Model: A1489) or a Samsung device (Model: GT-P5220), and their content was summarized by the lead author [18]. The apps were summarized into a tabular format to help assist participants during the semistructured interviews (Multimedia Appendix 1).

Textbox 1. Inclusion criteria of apps. (Please note that some selected apps were aimed at children and young people with attention-deficit hyperactivity disorder [ADHD], as well as the parents. These apps were included as long as they specifically stated they were for the parents as well. Apps were only excluded if the app was not intended for the parents in any way).

Inclusion criteria

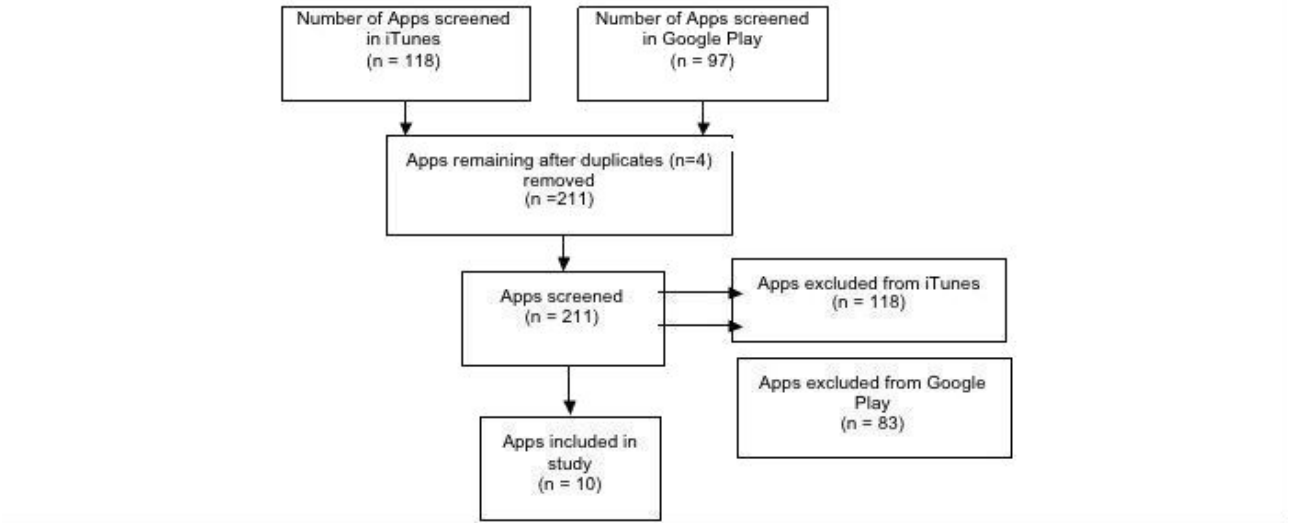
- An app that states it is aimed at a parent of a child or a young person with ADHD or attention deficit disorder
- Mobile app
- App is available in the English language

Textbox 2. Exclusion criteria of apps. (Please note that some selected apps were aimed at children and young people with attention-deficit hyperactivity disorder [ADHD], as well as the parents. These apps were included as long as they specifically stated they were for the parents as well. Apps were only excluded if the app was not intended for the parents in any way).

Exclusion criteria

- Does not state app is aimed at parent of a child or young people with ADHD
- Not a mobile app (eg, magazine)
- Not available in the English language
- Duplicate app

Figure 1. App selection process.



Participants

Convenience sampling was adopted to recruit specialist clinicians who work with children and young people with ADHD and their families, as well as parents of children and young people with ADHD. Parents were recruited via a Family Action group (a national charity that provides support for families) and an online ADHD community group. Clinicians were recruited via the NHS Child and Adolescent Mental Health Service in the South Yorkshire region. Participants were recruited until data saturation was achieved. Eligibility criteria for clinicians were that they had to be employed by a service that treats children and young people with ADHD.

Eligibility criteria for the parents of the children and young people with ADHD were as follows: (1) must be a parent of a young person who has a confirmed ADHD diagnosis and (2) must be able to provide details of the medication the child is prescribed.

Procedure

Ten separate semistructured interviews took place in the Sheffield region. Three interviews with parents involved interviewing parents jointly. Interviews lasted up to 60 min with the clinicians, and up to 90 min with the parents. The study received ethical approval from the University of Sheffield's School of Health and Related Research Ethics Committee (references 010768, 011377), as well as NHS Health Research Authority and Research and Development approval.

The researcher visited parents at the Family Action parent group and introduced the study. Parents then either contacted the researcher, or they requested that the researcher contacted them at a time convenient to them. Clinicians were approached via email. Interviews took place at a quiet convenient place with the clinicians and in the homes with the parents. Before the interviews, the clinicians and parents provided written informed consent.

Information about the study was sent to all participants at least 1 week before the interviews took place. At the beginning of each interview, the study was explained to the participants and questions were answered.

Clinicians and parents were presented with all 10 apps identified. Participants were given the opportunity to use the apps themselves during the interview. Interview discussions were guided by an interview schedule covering their views in four key areas:

- What makes a successful app?
- What makes an app less successful or unsuccessful?
- How could an app function benefit them as parents of young people with ADHD?
- How could apps for parents help manage ADHD or address difficulties in young people?

The two groups provided two unique perspectives: a user perspective and a clinician perspective. Participants also completed a short questionnaire on their demographic characteristics. Parents additionally provided details of their child's diagnosis and prescribed ADHD medication (where applicable).

Data Analysis

All interviews were audio-recorded and transcribed verbatim. Thematic analysis [51] was used to search for data patterns within and across the participant groups. LP and JP independently identified codes and themes from the transcripts. Discrepancies were resolved through group discussion in an iterative fashion between the authors. Themes identified aimed to capture the essence of the participant's views.

Table 1. Summary of how app developers describe the features of their apps.

iTunes apps			Android apps		
App	App claims	How apps meet their claims	App	App claims	How apps meet their claims
1	News updates and research about ADHD[a] and other conditions. Increases knowledge.	Condition-specific resource for news, features, and research	6	To help coordinate parents, caregivers, and teachers in the follow-up of children aged under 18 years with ADHD	Manage medication, plan daily activities, measure treatment results, self-assessment tools, and direct doctor and teacher communication via sister app
2	To change challenging daily routines into fun. Songs to help guide child to timely efficient task completion through consistency, repetition, rhythm, and rhyme.	Step-by-step directions of daily tasks through songs, monitoring chart, and coloring book reward	7	Addresses memory Provides information about ADHD	Memory games of different levels and ADHD key concepts quiz. Dialogues with a cartoon character. Links provided with ADHD information
3	Improves self-control, reduces hyperactivity, and improves attention, concentration, and focus.	Mindfulness Training	8	Monitor ADHD symptoms over time Download resources	Links to ADHD info or resources Sliders to record key times of day, charts, email charts, and appointment reminders
4	To help child improve dressing skills.	Learn to dress by dressing a cartoon character in order and imitate	9	Gain solutions to problem behaviors	Click on problem behavior and find a suggested solution and possible approaches. Related to special education
5	To help children and adults understand ADHD and how to manage it.	Interactive story about a boy and a character with ADHD	10	Create a visual schedule to support transition times during the daytime	User can illustrate sequence of tasks or substeps of a task and can label images with text

[a]ADHD: attention-deficit hyperactivity disorder.

Table 2. Demographic characteristics of parents.

Unique ID	Gender	Length of child's ADHD[a] diagnosis (years, months)	Child's age (years)	Other diagnoses of child	Child medicated?
P1	F[b]	Undergoing ADHD assessment	11	ASD[c]	No
P2	F	4, 0	15	Not applicable	Yes
P3	F	0, 7	13	ASD and anxiety	Yes
P4	F	0, 7	13	ASD and anxiety	Yes
P5	M	1, 10	10	ASD	Yes
P6	M	0, 8	9	ASD	No; under assessment
P7	M	0, 8	9	ASD	No; under assessment

[a]ADHD: attention-deficit hyperactivity disorder.

[b]F: female.

[c]ASD: autism spectrum disorder.

Table 3. Demographic characteristics of clinicians demonstrating a total of more than 57 years of experience working with children and young people with attention-deficit hyperactivity disorder (ADHD).

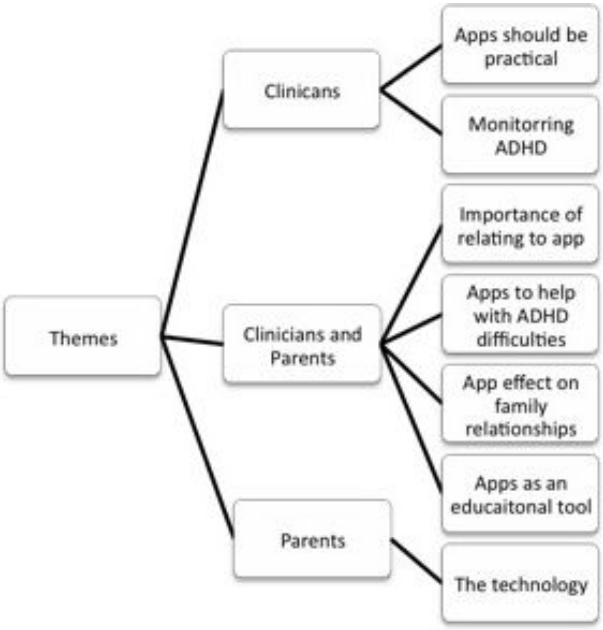
Unique ID	Gender	Length of time working with population (years)	Current job title
C[a]1	F[b]	6	Consultant child and adolescent psychiatrist
C2	F	12	Consultant child and adolescent psychiatrist
C3	F	0.5	Trainee 2 psychiatry
C4	F	10	Primary mental health worker
C5	M[c]	20	Consultant child and adolescent psychiatrist
C6	M	13	Consultant child and adolescent psychiatrist

[a]C: clinician.

[b]F: female.

[c]M: male.

Figure 2. Themes identified during thematic analysis.



Results

Five apps were identified from Google Play and five from iTunes in August 2016. Table 1 describes the claims the apps make within their individual descriptions and their contents.

Participant Characteristics

A total of 13 participants were recruited from February to March 2017 (clinicians) and October to November 2016 (parents). Participants included 6 clinicians and 7 parents of young people diagnosed with ADHD. Participant characteristics are reported in Table 2 (parents) and Table 3 (clinicians). Two of the 7 parents were recruited via an online ADHD parent group in South Yorkshire. Five of the 7 parents were recruited from an ADHD parent group in Sheffield. Parents are only referred to this group if their child was diagnosed or being assessed for ADHD. All parents were able to provide information regarding the ADHD medication prescribed to their children, when applicable.

Data was collected until data saturation was achieved. Transcripts were available for all 13 interviews. During analysis, agreement between the 2 primary coders was high. Seven themes were identified in total (Figure 2). Where similarities between parents and clinician views emerged, from the data, they are combined and discussed under the same themes. These themes are identified, compared, and discussed below with illustrative quotations.

Themes Identified by Parents and Clinicians

Importance of Relating to the App

Parents (n=7) and clinicians (n=4) believed it was important to relate to the app, especially when the app is aimed at both the parent and the child:

> *...if you could choose who you were, [referring to avatar in app] she [child] could relate a bit better.* [P1]

> *...that's an issue already [referring to gender and race of apps character], it's not even a boy is it, is it a girl or a boy?* [P1]

Clinicians (n=5) and parents (n=7) believed that apps should be visually attractive and have an appealing audio component (clinicians n=4; parents n=5):

> *I like quite visual things, I like visual things, I like stuff that takes 2 minutes, simple.* [P1]

Parents also believed that they should be able to alter apps to suit their own circumstances (n=7). Examples of this involve dressing a character in an order that is dictated by the parent; the parent could choose the items of clothing or the songs they play rather than the app dictating this. In terms of monitoring, they would like to choose the factors that they monitor rather than having these dictated by the app as well.

> *...having the option...at the beginning of being able to change the colors because he might go oh I can't put that on because I don't own a yellow t shirt...* [P2]

> *...like a [different] song every morning,...Michael Jackson, then Bob Marley, you know [tailoring the app to] what he [son] likes...if he woke up to something he likes it would put him in a better frame of mind [for the rest of the day]...* [P7]

Apps Should Target ADHD-Related Difficulties

Parents (n=7) and clinicians (n=6) also believed that the apps should target difficulties that specifically relate to ADHD, such as daily routines (clinicians n=3, parents n=7), behavior management (clinicians n=1), organizational skills (parents n=2), and to improve their communication with schools (clinicians n=2, parents n=2).

> *...if you've done that preparatory work [using an app to learn how to get dressed in the morning] with the child then hopefully that would make things easier [for the parents] in the mornings.* [health care professional 2, HCP2]

> *...they've got no organizational skills what so ever so that's [app targeting organizational skills] really good.* [P1]

Two parents acknowledged that apps have the potential to reduce their ADHD-related anxieties (parents n=2). One clinician (HCP1) stated how an app could do this and why it could be beneficial:

> *...the parents' reaction to the behavior...may be based on...day stress, my boss yelling at me so if I can do something to bring own my stress levels my response to the child with ADHD may be different.* [HCP1]

Almost all the participants (clinicians n=6, parents n=6) were able to highlight the benefits of using an app to monitor ADHD-related difficulties in ADHD:

> *...might spot trends as to why their behavior became the way it was.* [P2]

> *...to help you see where your kids struggling I would say.* [P4]

> *...good idea because...school can give more information...because I don't see my child with the medication so it would be useful if I could see what the teachers are saying...?* [P3]

One parent believed that an app shouldn't try to slow down a young person with ADHD but harness their hyperactivity by having an app that is also fast paced:

> *...having something that tries to slow em down just doesn't work. You're better off keeping them in one place by giving them something that will interact that's fast.* [P5]

App Effect on Family Relationships

Parents and clinicians (clinicians n=3, parents n=3) noted that apps could be used to improve relationships between the parent and their child:

> *...it's like more bonding time, education for myself and my son.* [P6]

> *I like the idea of parent and child working alongside [using an app].* [HCP1]

> *...bit of fun between the parent and child cos mornings and mealtimes and bedtimes can be really stressful so [the app] takes away some stress and introduces a bit of relationship building as well.* [HCP4]

It was also noted that apps could help address an ADHD diagnosis and what it means with siblings so that they can learn about their brother or sister's condition (clinicians n=2, parents n=3):

> *...I think it would be useful to use with siblings [to explain ADHD to them] as well an' they could interact with the game.* [HCP2]

Apps as an Educational Tool

Parents and clinicians liked the idea that apps could be used as an educational tool to learn about ADHD either via the app itself or via Web resources that are signposted within the apps (clinicians n=4, parents n=3). Parents also liked that some of the apps presented ADHD in a positive light (n=4):

> *...you can't keep going on courses for the rest of your life. And you can't always get it from a book as you either have to buy the book or the library doesn't have it.* [P5]

> *...being reminded of positive things I also good for you and the kids self-esteem...* [P3]

Clinicians acknowledged that apps should be culturally relevant (n=2). For example, one American app provided news stories that provided information that isn't always relevant to the United Kingdom:

> *...parents may be taking things on board that when you look into it there's not much evidence for. Also, it's promoting medications of therapies or whatever their whole healthcare systems structured differently to ours, which puts a bias on it...And also they have medication licenses that we don't.* [HCP2]

One clinician stated that app developers should be mindful that parents also have varying abilities:

> *...it's hard for parents to understand exactly what ADHD means...if you have a parent with...learning difficulties then its explained in an easy simple to understand way...* [HCP2]

Themes Identified by Clinicians

Apps Should Be Practical

Many of the clinicians stated that parents of children and young people with ADHD lead busy and often hectic lives. Therefore, apps should be simple to use and easily integrated into their daily routines if they are to use them routinely (n=4). One clinician stated that if apps are visually appealing, then this would contribute to the ease of use of the product, which is important. They also believed that many parents may not want to pay for apps (n=3):

> *I don't think the parents have got time for apps...I've met very few parents that will have time...most of them will have a sibling with ADHD or...younger children...* [HCP3]
>
> *...if you are if you're a parent…trying to catch up on something in a fairly fast way,...I want something that will be quick...* [HCP6]
>
> *I wouldn't spend that [as a parent].* [HCP6]

Apps to Monitor ADHD

Clinicians believed that apps could be a good tool to monitor the ADHD-related difficulties of a child, their symptoms relating to medication and also to provide reminders such as for appointments and taking ADHD medication (n=6):

> *...if they are able to keep that record [ADHD monitoring on an app] it may be helpful for the clinician who is monitoring the young person's care...* [HCP5]
>
> *...the best thing is...it will remind you say at 9 o'clock at night and you can record the day.* [HCP6; praising an app that reminds the parents to enter ADHD symptom information into the app on a daily basis]

Theme Identified by Parents

The Technology

Parents stressed the importance of the technology itself. They believed that an app isn't always necessary (n=5); they can be distracting, and accessing the technology can sometimes be problematic (n=5). They also discussed how some of the apps required the input of identifiable information. Parents raised concerns here regarding data security (n=3);

> *...that's all I do,...I type in ADHD in Twitter if I want to know about that I won't download an app for it.* [P1]
>
> *Once again the only thing is it's (the app) trapped in Apple world.* [P4]
>
> *...today your data might be safe, tomorrow your data maybe be going off on tour somewhere for a, for a price.* [P4]

Table 4 demonstrates the app characteristics identified by participants during semistructured interviews and which apps possessed these characteristics.

Discussion

Principal Findings

This study identified 10 apps that stated they were aimed at the parents of children and or young people with ADHD. Seven parents of young people with ADHD and 6 clinicians working with this population were interviewed. Interviews involved sharing the 10 identified apps with participants and asking what they think would make an app suitable for parents of children and young people with ADHD. Parents stated that technology access can be a barrier, that apps could enhance family relationships, apps have the potential to explain what ADHD means to effected siblings, apps should be flexible, and they could be used as an ADHD educational tool. Additionally, clinicians felt that apps would be useful for parents to monitor their child's ADHD symptoms and related difficulties and that apps should be practical, quick, and easy to use to account for the parents with often busy and chaotic lifestyles. The latter findings are consistent with the findings of Simons et al [40].

As living with and managing the ADHD with a child or young person can be incredibly stressful for parents, the National Institute for Health and Care Excellence guidelines recommend that all the parents of children and young people diagnosed with ADHD should be invited to attend parent management and support groups [19]. There are a number of possible parent training or management programs available [35-38]; however, they cannot provide ongoing support, reminders, or instant advice at a time of crisis.

We are living in a society where technology is routinely used and embedded into our lives. Many apps are advertised for parents of children and young people with ADHD, but there is little evidence to underpin them and little evidence to ascertain what makes such apps suitable for parents. The young people and clinicians whose views were sought in this study did not find that the apps reviewed fully met their expectations. The highest match was 5 out of 8; two apps scored 1 out of 8, and the mean score was 3.6. Parents wanted the apps to be visually pleasing and tailored to their own circumstances. Clinicians were keen to highlight the possibility of using apps to allow parents to monitor their child or young person's ADHD-related difficulties. Clinicians saw the opportunity to use this data clinically.

Table 4. Summary of what makes an app suitable for a parent of a child or young person with attention-deficit hyperactivity disorder (ADHD), according to the parents and the clinicians interviewed in this study and which apps identified in this study include these characteristics. The apps have been scored out of eight. The scores represent how each app fits with the views of the participants.

Characteristics identified by participants as likely to be positive	iTunes					Google Play				
	App 1	App 2	App 3	App 4	App 5	App 6	App 7	App 8	App 9	App 10
Visually pleasing (ie, includes bright colors)	No	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes
Allows personalization so that the user can relate to the app	No	No	No	No	No	Yes	No	Yes	No	Yes
Should help specifically with ADHD[a]-related difficulties	Aims to	Aims to	No	Aims to	Aims to	Aims to	Aims to	Aims to	No	No
Joint use (parent and child) to encourage health family relationships	No	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
Can be used as an educational tool or information source	Yes	No	No	No	Yes	No	Yes	No	Yes	No
App to monitor ADHD symptoms and difficulties	No	No	No	No	No	Yes	No	Yes	No	No
Should be practical (easy to use or embed into daily routine)	Yes	No	No	No	Yes	No	Yes	Yes	No	Yes
Attempts to improve daily routines	No	Yes	No	Yes	No	No	No	No	No	Yes
App score (out of 8)	3	4	1	4	5	3	5	5	1	5

[a]ADHD: attention-deficit hyperactivity disorder.

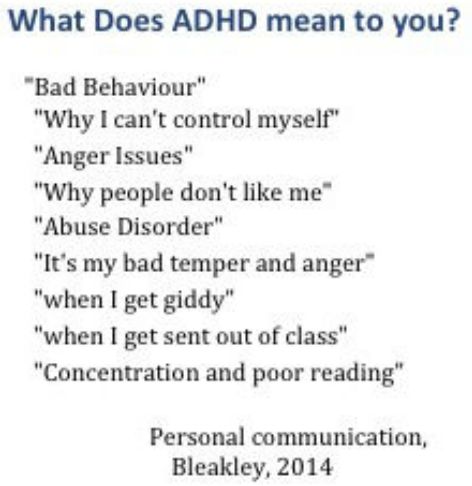
Increasing knowledge about a disorder, for example, by using psychoeducation, is recognized as important in improving management and then outcomes in chronic disorders. The knowledge of parents or carers of children with ADHD and, in particular, of young people with ADHD themselves, has been shown to be poor [47]. When young people in an ADHD clinic were asked what ADHD meant to them, they replied with negative answers that did not accurately represent the clinician's views, as noted by Bleakley C (2014). A summary of these views can be found in Figure 3. Although young people valued direct information sharing with their doctor or nurse, they were also comfortable with using technology to improve their understanding [47].

The perspectives of both the clinicians and the parents are needed to build up a full picture. Clinicians are keen to make use of apps to promote positive behaviors and target specific needs. The views of the parents give guidance as to whether they would actually use the app and whether they felt it was likely to be helpful.

A Canadian study [52] considered parents' attitudes to medical technology (genetic studies and magnetic resonance imaging scanning) and concluded that parents were in fact "In search of Anything that would help." Indeed, their interviews reflected parents' views that their needs were not being met despite increased understanding among health and education professionals. The choice of medications for ADHD has increased in recent years, but there is still a reluctance to use medication, and for many young people, achieving symptom improvement throughout their waking times, remains a challenge. Technology can now be accessed at any time in most places by most people, so it may offer the much needed extra support to young people and their families.

If technology is to be reliably helpful, some quality standards must be met. Apps are often developed rapidly with little evidence base. The pressure is often for financial success rather than proven benefit. This study gives us some insight into the quality of currently available apps and suggests that more research is needed to harness technology to benefit young people with ADHD and their families.

Figure 3. Replies from teenagers in an attention-deficit hyperactivity disorder (ADHD) clinic.



Conclusions

This research suggests that currently the top 10 listed apps marketed for parents with young people with ADHD do not seem likely to meet their needs and crucially, do not have the key components of they believe will make an app helpful.

According to our participants, key components of a successful app are likely to be; pleasing visuals, the facility to personalize, the ability to help with specific ADHD difficulties, design for joint use by a parent and young person, use as an ADHD educational tool, use in monitoring ADHD-related difficulties and symptoms, easy to use, possible use to improve daily routines.

Future research is needed on the value of apps for parents of children and young people with ADHD and, in particular, any positive role for apps in the management of ADHD in this age group.

Future research is needed to develop apps in coproduction with key stakeholders and explore the effectiveness of technology-based interventions.

This could result in setting standards for apps for ADHD and other chronic conditions, which could guide users to the optimal apps for their needs.

Acknowledgments

This research was funded and supported by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care Yorkshire and Humber. The views and opinions expressed are those of the authors and are not necessarily those of the National Health Service, the National Institute for Health Research, or the UK's Department of Health.

This study was funded by the University of Sheffield's Teaching Assistant Sheffield undergraduate Research Experience Scheme and the Engineering and Physical Sciences Research Council Vacation Bursary Scheme.

The authors would like to thank the children and young people and their families and the clinicians for taking part and sharing their expertise.

Conflicts of Interest

None declared.

Multimedia Appendix 1

App summary.

[PDF File (Adobe PDF File), 18KB - mhealth_v5i10e149_app1.pdf]

References

1. Biederman J. Attention-deficit/hyperactivity disorder: a selective overview. Biol Psychiatry 2005;57(11):1215-1220. [doi: 10.1016/j.biopsych.2004.10.020] [Medline: 15949990]
2. Faraone SV, Biederman VJ, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. Psychol Med 2006;36(02):159-165. [doi: 10.1017/S003329170500471X] [Medline: 16420712]

3. Harpin VA. The effect of ADHD on the life of an individual, their family, and community from preschool to adult life. Arch Dis Child 2005;90(suppl 1):i2-i7. [doi: 10.1136/adc.2004.059006]
4. Biederman J, Newcorn J, Sprich S. Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. Am J Psychiatry 1991;148(5):564-577 [FREE Full text]
5. Kadesjö B, Gillberg C. The comorbidity of ADHD in the general population of Swedish school-age children. J Child Psychol Psychiatry 2001 May;42(4):487-492. [Medline: 11383964]
6. Yoshimasu K, Barbaresi WJ, Colligan RC, Voigt RG, Killian JM, Weaver AL, et al. Childhood ADHD is strongly associated with a broad range of psychiatric disorders during adolescence: a population-based birth cohort study. J Child Psychol Psychiatry 2012 Oct;53(10):1036-1043 [FREE Full text] [doi: 10.1111/j.1469-7610.2012.02567.x] [Medline: 22647074]
7. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry 2007 Jun;164(6):942-948. [doi: 10.1176/ajp.2007.164.6.942] [Medline: 17541055]
8. Barkley RA, Fischer M, Edelbrock CS, Smallish L. The adolescent outcome of hyperactive children diagnosed by research criteria: I. An 8-year prospective follow-up study. J Am Acad Child Adolesc Psychiatry 1990 Jul;29(4):546-557. [doi: 10.1097/00004583-199007000-00007] [Medline: 2387789]
9. Gittelman R, Mannuzza S, Shenker R, Bonagura N. Hyperactive boys almost grown up. I. Psychiatric status. Arch Gen Psychiatry 1985 Oct;42(10):937-947. [Medline: 4037987]
10. Weiss G, Hechtman L, Milroy T, Perlman T. Psychiatric status of hyperactives as adults: a controlled prospective 15-year follow-up of 63 hyperactive children. J Am Acad Child Psychiatry 1985 Mar;24(2):211-220. [Medline: 3989165]
11. Faraone SV. Attention deficit hyperactivity disorder and premature death. Lancet 2015 May 30;385(9983):2132-2133. [doi: 10.1016/S0140-6736(14)61822-5] [Medline: 25726517]
12. Hiscott R. Journals.lww.: Neurology Now; 2015. ADHD Alert: Risk of Premature Death Warrants Early Treatment URL: http://journals.lww.com/neurologynow/blog/breakingnews/Pages/post.aspx?PostID=266 [accessed 2017-09-27] [WebCite Cache ID 6tmy9XqQH]
13. Young J. ADHD is a risk factor for premature death, Danish study shows. Br Med J 2015 Feb 25;350:h1094. [Medline: 25721134]
14. Biederman J, Monuteaux MC, Mick E, Spencer T, Wilens TE, Silva JM, et al. Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. Psychol Med 2006 Feb;36(2):167-179. [doi: 10.1017/S0033291705006410] [Medline: 16420713]
15. Mannuzza S, Klein RG, Bessler A, Malloy P, LaPadula M. Adult psychiatric status of hyperactive boys grown up. Am J Psychiatry 1998 Apr;155(4):493-498. [doi: 10.1176/ajp.155.4.493] [Medline: 9545994]
16. Taylor E, Chadwick O, Heptinstall E, Danckaerts M. Hyperactivity and conduct problems as risk factors for adolescent development. J Am Acad Child Adolesc Psychiatry 1996 Sep;35(9):1213-1226. [doi: 10.1097/00004583-199609000-00019] [Medline: 8824065]
17. Telford C, Green C, Logan S, Langley K, Thapar A, Ford T. Estimating the costs of ongoing care for adolescents with attention-deficit hyperactivity disorder. Soc Psychiatry Psychiatr Epidemiol 2013 Feb;48(2):337-344. [doi: 10.1007/s00127-012-0530-9] [Medline: 22699685]
18. Doshi JA, Hodgkins P, Kahle J, Sikirica V, Cangelosi MJ, Setyawan J, et al. Economic impact of childhood and adult attention-deficit/hyperactivity disorder in the United States. J Am Acad Child Adolesc Psychiatry 2012 Oct;51(10):990-1002.e2. [doi: 10.1016/j.jaac.2012.07.008] [Medline: 23021476]
19. National Institute for Health and Care Excellence. NICE.: National Institute for Health and Care Excellence; 2013. Attention deficit hyperactivity disorder: NICE quality standard [QS39] URL: https://www.nice.org.uk/guidance/qs39 [accessed 2017-09-27] [WebCite Cache ID 6tmya9aHL]
20. Parker J, Wales G, Chalhoub N, Harpin V. The long-term outcomes of interventions for the management of attention-deficit hyperactivity disorder in children and adolescents: a systematic review of randomized controlled trials. Psychol Res Behav Manag 2013;6:87-99. [doi: 10.2147/PRBM.S49114]
21. Visser SN, Danielson ML, Bitsko RH, Holbrook JR, Kogan MD, Ghandour RM, et al. Trends in the parent-report of health care provider-diagnosed and medicated attention-deficit/hyperactivity disorder: United States, 2003-2011. J Am Acad Child Adolesc Psychiatry 2014 Jan;53(1):34-46.e2 [FREE Full text] [doi: 10.1016/j.jaac.2013.09.001] [Medline: 24342384]
22. Charach A, Yeung E, Volpe T, Goodale T, dosReis S. Exploring stimulant treatment in ADHD: narratives of young adolescents and their parents. BMC psychiatry 2014;14(1):1. [doi: 10.1186/1471-244X-14-110]
23. Miller VA, Drotar D. Decision-making competence and adherence to treatment in adolescents with diabetes. J Pediatr Psychol 2007 Mar;32(2):178-188. [doi: 10.1093/jpepsy/jsj122] [Medline: 16717139]
24. Edlund MJ, Wang PS, Berglund PA, Katz SJ, Lin E, Kessler RC. Dropping out of mental health treatment: patterns and predictors among epidemiological survey respondents in the United States and Ontario. Am J Psychiatry 2002 May;159(5):845-851. [doi: 10.1176/appi.ajp.159.5.845] [Medline: 11986140]
25. Greeno C, Anderson CM, Shear K, Mike G. Initial treatment engagement in a rural community mental health center. Psychiatr Serv 1999 Dec;50(12):1634-1636. [doi: 10.1176/ps.50.12.1634] [Medline: 10577887]

26. Hamilton W, Round A, Sharp D. Patient, hospital, and general practitioner characteristics associated with non-attendance: a cohort study. Br J Gen Pract 2002 Apr;52(477):317-319 [FREE Full text] [Medline: 11942451]
27. O'Brien A, Fahmy R, Singh SP. Disengagement from mental health services. A literature review. Soc Psychiatry Psychiatr Epidemiol 2009 Jul;44(7):558-568. [doi: 10.1007/s00127-008-0476-0] [Medline: 19037573]
28. Wang J. Mental health treatment dropout and its correlates in a general population sample. Med Care 2007 Mar;45(3):224-229. [doi: 10.1097/01.mlr.0000244506.86885.a5] [Medline: 17304079]
29. Pliszka SR. Comorbidity of attention-deficit/hyperactivity disorder with psychiatric disorder: an overview. J Clin Psychiatry 1998;59(Suppl 7):50-58 [FREE Full text] [Medline: 9680053]
30. Loe IM, Feldman HM. Academic and educational outcomes of children with ADHD. J Pediatr Psychol 2007 Jul;32(6):643-654. [doi: 10.1093/jpepsy/jsl054] [Medline: 17569716]
31. Frazier TW, Youngstrom EA, Glutting JJ, Watkins MW. ADHD and achievement: meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. J Learn Disabil 2007;40(1):49-65. [doi: 10.1177/00222194070400010401] [Medline: 17274547]
32. McGee R, Partridge F, Williams S, Silva PA. A twelve-year follow-up of preschool hyperactive children. J Am Acad Child Adolesc Psychiatry 1991 Mar;30(2):224-232. [doi: 10.1097/00004583-199103000-00010] [Medline: 2016226]
33. Smith E, Koerting J, Latter S, Knowles MM, McCann DC, Thompson M, et al. Overcoming barriers to effective early parenting interventions for attention-deficit hyperactivity disorder (ADHD): parent and practitioner views. Child Care Health Dev 2015 Jan;41(1):93-102 [FREE Full text] [doi: 10.1111/cch.12146] [Medline: 24814640]
34. Barkley RA, Fischer M, Smallish L, Fletcher K. The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. J Abnorm Psychol 2002 May;111(2):279-289. [Medline: 12003449]
35. Loren RE, Vaughn AJ, Langberg JM, Cyran JE, Proano-Raps T, Smolyansky BH, et al. Effects of an 8-session behavioral parent training group for parents of children with ADHD on child impairment and parenting confidence. J Atten Disord 2015 Feb;19(2):158-166 [FREE Full text] [doi: 10.1177/1087054713484175] [Medline: 23599209]
36. Thompson MJJ, Laver-Bradbury C, Ayres M, Le PE, Mead S, Dodds C, et al. A small-scale randomized controlled trial of the revised new forest parenting programme for preschoolers with attention deficit hyperactivity disorder. Eur Child Adolesc Psychiatry 2009 Oct;18(10):605-616. [doi: 10.1007/s00787-009-0020-0] [Medline: 19404717]
37. Porzig-Drummond R, Stevenson RJ, Stevenson C. The 1-2-3 Magic parenting program and its effect on child problem behaviors and dysfunctional parenting: a randomized controlled trial. Behav Res Ther 2014 Jul;58:52-64. [doi: 10.1016/j.brat.2014.05.004] [Medline: 24929928]
38. Sanders MR, Kirby JN, Tellegen C, Day JJ. The Triple P-Positive Parenting Program: a systematic review and meta-analysis of a multi-level system of parenting support. Clin Psychol Rev 2014 Jun;34(4):337-357. [doi: 10.1016/j.cpr.2014.04.003] [Medline: 24842549]
39. Livingstone S, Haddon L, Görzig A, Ólafsson K. LSE. 2011. Risks and safety for children on the internet: the UK report: full findings from the EU Kids Online survey of UK 9-16 year olds and their parents URL: http://www.lse.ac.uk/media@lse/research/EUKidsOnline/EU%20Kids%20II%20(2009-11)/National%20reports/UKReport.pdf [accessed 2017-09-29] [WebCite Cache ID 6tpyjVdUl]
40. Simons L, Valentine AZ, Falconer CJ, Groom M, Daley D, Craven MP, et al. Developing mHealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. JMIR Mhealth Uhealth 2016 Mar 23;4(1):e31 [FREE Full text] [doi: 10.2196/mhealth.5009] [Medline: 27009498]
41. Spachos D, Chifari A, Chiazzese G, Merlo G, Doherty G, Bamidis P. WHAAM: A mobile application for ubiquitous monitoring of ADHD behaviors. 2014 Presented at: Interactive Mobile Communication Technologies and Learning (IMCL), 2014 International Conference on; 13-14 November, 2014; Greece. [doi: 10.1109/IMCTL.2014.7011153]
42. Pina L, Rowan K, Roseway A, Johns P, Hayes GR, Czerwinski M. In situ cues for ADHD parenting strategies using mobile technology. 2014 Presented at: PervasiveHealth '14 Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare; May 20 - 23, 2014; Germany. [doi: 10.4108/icst.pervasivehealth.2014.254958]
43. Sonne T, Obel C, Grønbæk K. Designing Real Time Assistive Technologies: A Study of Children with ADHD. 2015 Presented at: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction; December 07 - 10, 2015; Australia p. 34-38. [doi: 10.1145/2838739.2838815]
44. Young Z, Craven MP, Groom M, Crowe J. Snappy App: a mobile continuous performance test with physical activity measurement for assessing Attention Deficit Hyperactivity Disorder. 2014 Presented at: International Conference on Human-Computer Interaction; 2014; Greece p. 363-373. [doi: 10.1007/978-3-319-07227-2_35]
45. British Standards Institution. Innovate UK Technology Strategy Board. 2015. Health and wellness apps - Quality criteria across the life cycle - Code of practice URL: https://shop.bsigroup.com/upload/271432/PAS%20277%20(2015)bookmarked.pdf [accessed 2017-09-29] [WebCite Cache ID 6tpz6zP71]
46. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 07;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]

47. Bussing R, Zima BT, Mason DM, Meyer JM, White K, Garvan CW. ADHD knowledge, perceptions, and information sources: perspectives from a community sample of adolescents and their parents. J Adolesc Health 2012 Dec;51(6):593-600 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.03.004] [Medline: 23174470]
48. Cailean I. Trademob. 2015. Extensive guide to app store optimization (ASO) in 2015? part 1: apple app store URL: http://www.trademob.com/app-store-optimization-guide-apple/ [accessed 2017-09-27] [WebCite Cache ID 6tn0aOi99]
49. Cailean I. Trademob. 2015. Extensive guide to app store optimization (ASO) in 2015? part 2: google play store URL: http://www.trademob.com/app-store-optimization-guide-google/ [accessed 2017-01-19] [WebCite Cache ID 6ndTrjaCK]
50. Joddrell P, Hernandez A, Astell AJ. Identifying Existing, Accessible Touchscreen Games for People Living with Dementia. 2016 Presented at: International Conference on Computers Helping People with Special Needs; 2016; Austria.
51. Pope C, Ziebland S, Mays N. Analysing qualitative data. Br Med J 2000;320(7227):114-116. [doi: 10.1136/bmj.320.7227.114]
52. Borgelt EL, Buchman DZ, Weiss M, Illes J. In search of "anything that would help" parent perspectives on emerging neurotechnologies. J Atten Disord 2014 Jul;18(5):395-401. [doi: 10.1177/1087054712445781] [Medline: 22617864]

Abbreviations

ADHD: attention-deficit hyperactivity disorder
ASD: autism spectrum disorder
ASO: app store optimization
F: Female
HCP: health care professional
M: Male
NHS: National Health Service

Edited by G Eysenbach; submitted 28.04.17; peer-reviewed by C Hollis, Y Liu; comments to author 11.08.17; revised version received 17.08.17; accepted 06.09.17; published 13.10.17.

Please cite as:
Powell L, Parker J, Harpin V
ADHD: Is There an App for That? A Suitability Assessment of Apps for the Parents of Children and Young People With ADHD
JMIR Mhealth Uhealth 2017;5(10):e149
URL: http://mhealth.jmir.org/2017/10/e149/
doi:10.2196/mhealth.7941
PMID:29030325

Original Paper

Commercially Available Smartphone Apps to Support Postoperative Pain Self-Management: Scoping Review

Chitra Lalloo[1], PhD; Ushma Shah[2], DNB, EDRA, EDAIC, FRCA; Kathryn A Birnie[1,3], PhD; Cleo Davies-Chalmers[1]; Jordan Rivera[1], BScN; Jennifer Stinson[1,3,4], RN, PhD, CPNP; Fiona Campbell[4,5], MD, FRCA

[1]The Hospital for Sick Children, Department of Child Health Evaluative Sciences, Toronto, ON, Canada

[2]Department of Anesthesiology and Perioperative Medicine, University of Western Ontario, London, ON, Canada

[3]Lawrence S. Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada

[4]Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, Toronto, ON, Canada

[5]Department of Anesthesia, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Chitra Lalloo, PhD
The Hospital for Sick Children
Department of Child Health Evaluative Sciences
686 Bay Street
Toronto, ON, M5G1X8
Canada
Phone: 1 4168137654 ext 302332
Fax: 1 4168138501
Email: chitra.lalloo@sickkids.ca

Abstract

Background: Recently, the use of smartphones to deliver health-related content has experienced rapid growth, with more than 165,000 mobile health (mHealth) apps currently available in the digital marketplace. With 3 out of 4 Canadians currently owning a smartphone, mHealth apps offer opportunities to deliver accessible health-related knowledge and support. Many individuals experience pain after surgery, which can negatively impact their health-related quality of life, including sleep, emotional, and social functioning. Smartphone apps that provide remote real-time monitoring and symptom management have the potential to improve self-management skills in patients experiencing postoperative pain. Increased confidence and practice of self-management skills could contribute to decreased postoperative pain and reduce risk of developing persistent pain. Published reviews of general pain self-management apps demonstrate a lack of evidence-based content, theoretical grounding, and health care professional involvement. However, no review to date has focused on the app marketplace specific for individuals with postoperative pain.

Objective: The aim of this study was to characterize and critically appraise the content and functionality of commercially available postoperative pain self-management apps.

Methods: An electronic search and extraction was conducted between December 2016 and March 2017 of the official Canadian app stores for the three major smartphone operating systems (iPhone operating system [iOS], Android, and Windows). Stores were searched separately using predetermined search terms. Two authors screened apps based on information provided in the public app description. Metadata from all included apps were abstracted into a standard spreadsheet. Two authors verified the data with reference to the apps and downloaded apps themselves. The content and functionality of each app as it pertained to postoperative pain self-management was rated.

Results: A total of 10 apps met the inclusion criteria. All included apps were designed exclusively for the Android platform. Education was the most common self-management feature offered (8/10, 80%), with none of the apps offering features related to goal setting or social support. Overall, no single app was comprehensive in terms of pain self-management content. Five (50%) apps reported the involvement of a health care provider in their development. However, not a single app involved end users in their development, and none of the apps underwent scientific evaluation. Additionally, none of the apps were designed for use in pediatric patients.

Conclusions: Currently available postoperative pain apps for patients lack evidence-based content, goal setting, and social support functions. There is a need to develop and test comprehensive theory-based apps to support patients with pain self-management care following surgery.

(JMIR Mhealth Uhealth 2017;5(10):e162) doi:10.2196/mhealth.8230

KEYWORDS

pain, postoperative; smartphone; mobile applications; review; pain management; self care

Introduction

Background

More than 80% of patients experience acute postoperative pain after surgery, of whom approximately 75% report their pain intensity as moderate, severe, or extreme [1,2]. Unrelieved or undertreated acute postoperative pain can delay remobilization, lead to increased opioid use and related side effects, and negatively impact all aspects of health-related quality of life [3-5]. Greater confidence in one's ability to control pain is referred to as pain coping self-efficacy. Improvements in pain coping self-efficacy may contribute to decreased postoperative pain and reduced risk for developing persistent postoperative pain, an expensive and debilitating health problem [6,7]. Thus, the prevention or minimization of postoperative pain is critically important.

Remote real-time monitoring and symptom management support for postoperative pain patients is a relatively new notion made possible by technological advances. Smartphones, for instance, are now widely used across the life span [8,9]. Smartphones have the potential to empower people with postoperative pain who are at increased risk for undertreatment of pain and disruption in activities of daily living [10]. Smartphones can support self-management by (1) improving self-monitoring of pain and other symptoms in everyday environments (eg, home, hospital settings), (2) promoting pain self-efficacy and appropriate self-care (eg, adherence to prescribed medications, cognitive-behavioral pain coping strategies), and (3) minimize barriers to optimal pain treatment (eg, lack of transportation to tertiary care centers) [11,12].

Existing Reviews of Pain Management Apps

There are a growing number of generalized pain self-management smartphone apps available for users to download on their personal mobile devices. In a 2011 scoping review, Rosser and Eccleston identified 111 such apps [13]. A majority were designed for the iPhone (iPhone operating system, iOS) platform, with fewer being available for Android or BlackBerry operating systems. The primary app functions included education skills training (50.5%; 56/111), self-monitoring (eg, pain diary, 26.1%; 29/111), and relaxation training (21.6%; 24/111). Importantly, 85.6% (95/111) of the identified apps did not report the involvement of health care professionals in their design or evaluation of content. None of the apps were focused on postoperative pain [13]. In 2012, Reynoldson and colleagues conducted an updated search of the iOS and Android stores [14]. This search only included apps with a primary self-monitoring function (ie, allowing patients to track their pain episodes). Apps were excluded if their intended use was limited to pain from specific conditions such as arthritis, inflammatory bowel disease, or migraine. The authors identified 12 eligible apps, which then underwent content analysis and usability testing. Their review identified a lack of user and clinician engagement in pain app development, as well as variation in app quality [14]. In 2015, Lalloo and colleagues conducted an updated scoping review of available pain management apps [15]. A total of 279 apps met the inclusion criteria. Pain self-care skill support was the most common self-management function (77.4%; 216/279). Apps also purported providing patients with the ability to engage in pain education (45.9%; 128/279), self-monitoring (19.0%; 53/279), social support (3.6%; 10/279), and goal setting (0.7%; 2/279). No apps were comprehensive in terms of pain self-management, with the majority of apps including only a single self-management function (58.8%; 164/279). Additionally, only 8.2% (23/279) of the apps included a health care professional in their development, not a single app provided a theoretical rationale, and only 1 app underwent scientific evaluation [15].

Objectives

To our knowledge, no scoping review to date has focused on apps for postoperative pain management. In addition, no scoping reviews to date have sought to identify apps for both pediatric and adult users. Therefore, the objectives of this study were to (1) characterize the current field of patient-focused pain self-management apps across all major smartphone platforms in Canada; (2) critically appraise the content and functionality of these apps, including self-care skill support, pain education, self-monitoring, social support, and goal setting; and (3) identify gaps in the field to guide future development of an app to support individuals with postoperative pain.

Methods

An electronic search and extraction was conducted between December 2016 and March 2017 of the official Canadian app stores for the 3 major smartphone operating systems: iOS (iTunes App Store), Android (Google Play Store), and Windows (Windows Store). The entire stores were searched separately using predetermined search terms (ie, no restrictions related to store subcategories such as "health and wellness" were imposed). The search terms were as follows: "Pain, Surgery," "Pain, Post-Surgery," "Pain, Post-Surgical," "Pain, Operative," "Pain, Operation," "Pain Management, Surgery," "Pain Management, Post-Surgery," "Pain Management, Post-Surgical," "Pain Management, Operative," "Pain Management, Operation."

No limits were imposed related to language or date of app publication. Given that the process for app indexing exhibited variability across app stores, a calibration exercise was conducted before app selection in order to verify our ability to detect postsurgical pain apps. Specifically, we tested our search criteria for the ability to find postsurgical pain apps that were known to be currently available in the app stores. In all cases, our search criteria identified the apps we expected to find.

Screening and Selection of Apps

Apps were included in the study if the primary intended user of the app was a person undergoing surgery, and a stated goal of the app was to provide education, tools, or advice related to managing pain after surgery. Apps were excluded if they focused only on the services offered by specific hospitals because they were intended as advertisements for these sites. These *clinic apps* typically included only lists of the pain management services and personnel available at the hospital and directions to the clinic rather than self-management programming. Apps were also excluded if they were classified as *e-books* by the respective app store or were judged by the reviewers as such. An *e-book* was defined as an app that did not provide any additional content or functionality beyond a textbook (eg, written content identical to a book). Authors performed app selection independently and all discrepancies regarding selection were resolved through discussion with a third author. Apps were screened based on the information provided in the *app description* section of each respective app store. Apps that met all screening criteria were downloaded and reviewed to confirm eligibility, and then data abstraction and content assessment were completed.

Data Abstraction

Metadata from all included apps were abstracted into a Microsoft Excel spreadsheet. Abstracted metadata included app name, URL, app description, customer rating on the day of store search, number of installations, and price. A systematic approach to data abstraction was used. Specifically, 2 authors (CDC and JR) abstracted all data and 2 others (CL and US) verified the data with reference to the app website and the downloaded apps themselves.

Content Assessment

To assess the comprehensiveness of app self-management content, the following criteria were used: (1) having a postoperative pain tracking function, (2) ability to set goals related to improving pain and functioning, (3) provision of skills training related to specific pain self-care strategies, (4) provision of social support, (5) provision of education related to surgery, and (6) provision of education related to pain after surgery. For each criterion, we assessed coverage as either *present* or *absent*. Similar criteria were used in our 2015 published scoping review [15] of general pain self-management apps. This list is based on the features of pain self-management programs that have been examined for effectiveness in scientific trials [16-19].

We also assessed whether a regulated health care professional had provided clinical expertise related to development of the app content and function. An app was required to reference the involvement of a health care professional listed in the Ontario Regulated Health Professions Act [20] to meet this criterion. This list contains 26 health care regulatory colleges to whom the Ontario government has endowed the capacity to regulate the practices of corresponding health care professionals to ensure safe and ethical patient care. The list includes professions such as physicians, nurses, and physiotherapists. If an app claimed the input of a profession that was not listed in the Act in its development, we assessed the health care professional involvement as *absent*. The involvement of end users (eg, surgery patients) in app development was also documented as *present* or *absent*. Apps were reviewed to determine whether they were designed for pediatric (aged 18 years or younger) or adult users.

The app store descriptions were also screened for any reference to the app being part of formal scientific research. The app description and corresponding developer website (if available) were reviewed to determine whether any specific theoretical framework had been used to guide the development of app content. Finally, large publicly accessible scientific literature databases (ie, NCBI PubMed and Google Scholar) were searched using the app name as a query for any published research related to the app. The MyHealthApps website was also reviewed as per the methods of de la Vega and Miro [21]. Descriptive statistics were used to summarize the results of the content assessment.

Results

App Screening

Our search strategy across iTunes, Windows, and Google Play app stores identified a total of 1019 apps (923 Android, 93 iPhone, and 3 Windows), which were screened according to the inclusion and exclusion criteria. Only 10/1019 apps (0.98% of total screened) met all inclusion criteria and were included in the analysis (see Figure 1 for PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses] flowchart) [22]. All included apps were designed exclusively for the Android platform (see Table 1 for included apps).

Self-Management Content of Reviewed Apps

Education related to postoperative pain was the most common self-management feature offered across the identified apps. This educational content largely focused on self-care advice following surgery, particularly physiotherapy exercises to support return to mobility. However, there was no specific content related to emotional factors such as anxiety or catastrophizing related to surgery. As well, none of the apps offered features related to goal setting or social support. Overall, no single app was comprehensive in terms of pain self-management content. Additionally, no apps were designed specifically for pediatric patients. See Figure 2 for a summary of the self-management components across all reviewed apps.

Table 1. Postoperative pain apps included in the review.

#	App	Developer
1	Activity Heals 1on1	Elevenity
2	Breast Implant Exercises	Chicago Plastic Surgery
3	BW MyTENS	Visiomed Lab
4	Doado Your Health Companion	Alex Spriet
5	Fuse—Post Op. Journal	Imbas Solutions
6	Healing Power	Phil Shapiro
7	My Hip Rehab	App Squid
8	Physiotherapy Help Guide	Creativity Knowledge App
9	Pocket Physio	Care UK HC
10	Surgeon on Call	Phoenix Medical Consulting

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of app review process.

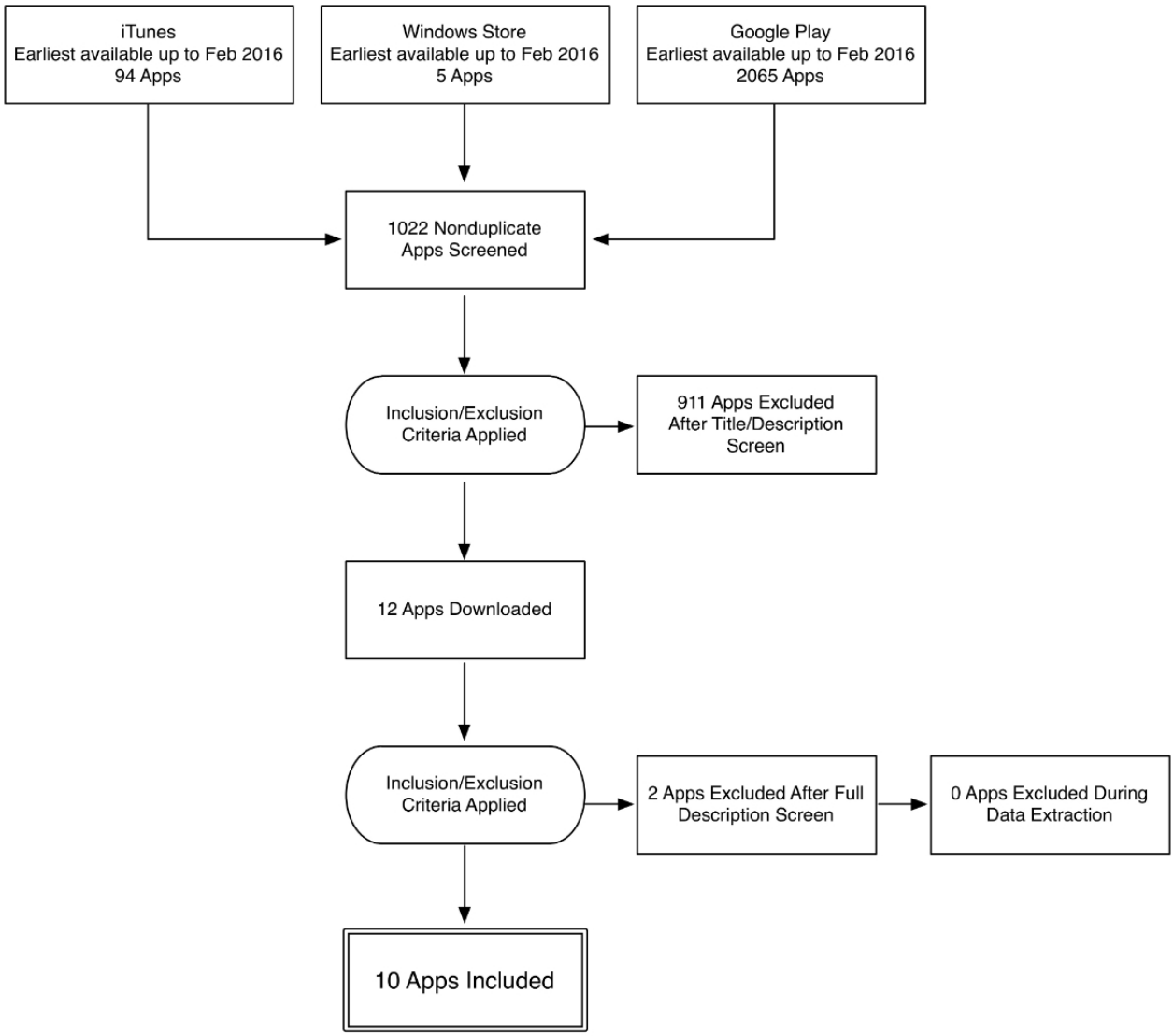


Figure 2. Self-management content across reviewed apps.

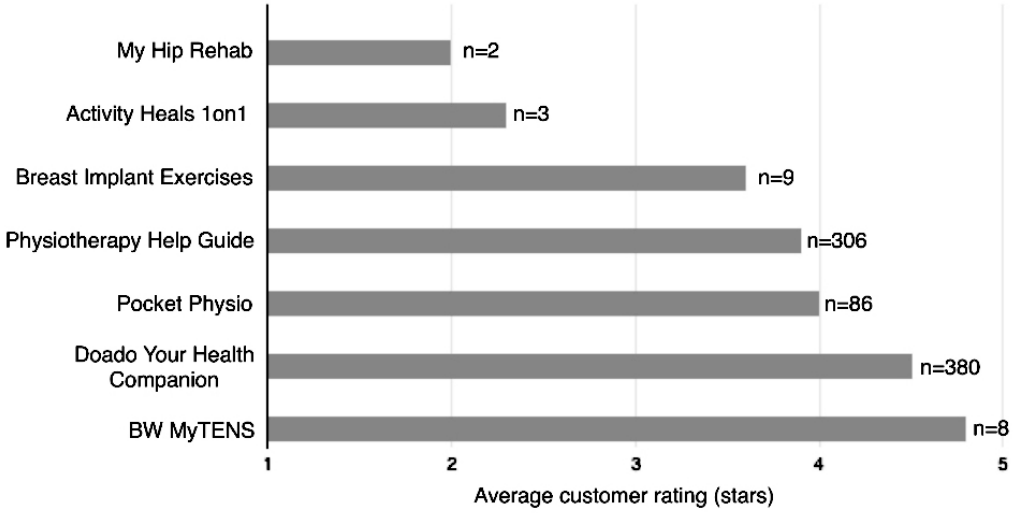
App Cost, Customer Ratings, and Number of Installs

Most apps (8/10, 80%) were free to download with the exceptions of *Surgeon on Call* and *My Hip Rehab*, which were associated with charges of Can $0.99 and Can $10.22, respectively. App store customers had the option of rating each app on a scale ranging from 1 to 5 stars (no anchors provided). There was a high variability in the number of users who chose to rate each app (range of 2-380 raters per app). As seen in Figure 3, the average customer rating ranged from 2 to 4.8 stars. However, fewer than 10 customers reviewed most (7/10, 70%) of the apps. The number of app installs ranged from as low as 10 to as high as 100,000 across apps (see Table 2).

Table 2. App installations. Categories provided by Android Google Play Store. Data collected on April 26, 2017.

Number of installs	Apps
10-50	Surgeon on Call; My Hip Rehab
100-500	Fuse–Post Op. Journal; Healing Power
500-1000	Activity Heals 1on1; BW MyTENS
1000-5000	Breast Implant Exercises
5000-10000	Pocket Physio
10,000-50,000	Doado Your Health Companion
50,000-100,000	Physiotherapy Help Guide

Figure 3. Customer-generated app ratings. *No customer ratings, or less than 2 ratings, were available for three apps: Fuse – Post-Op Journal, Surgeon on Call, and Healing Power.



Involvement of Health Care Providers, End Users, and Scientific Evaluation

Overall, 5/10 (50%) apps reported the involvement of a health care provider in their development. In summary, *Doado Your Health Companion* and *Pocket Physio* involved a physiotherapist, while *Breast Implant Exercises* and *Healing Power* involved a physician, and *Surgeon on Call* involved a specialist (surgeon). The other apps did not report any involvement of health care professionals in their development. None of the apps reported involvement of end users (eg, surgical patients) in their development. No evidence was found indicating that any of the apps had been scientifically evaluated.

Discussion

Principal Findings

This review demonstrates the paucity of high-quality apps that are commercially available to support self-management of postoperative pain for either adult or pediatric users. All included apps were designed exclusively for the Android platform. Education was the most common self-management feature offered (8/10, 80%), while none of the apps offered features related to goal setting or social support. Overall, no single app was comprehensive in terms of pain self-management content. Five (5/10, 50%) apps reported the involvement of a health care provider in their development. However, not a single app involved end users in their development and none of the apps underwent scientific evaluation. Additionally, none of the apps were designed for use in pediatric patients.

Customer ratings for the apps were moderate on average (mean 3.6, standard deviation 1.1). However, it is important to highlight that very few users chose to provide an app rating. For instance, while *Pocket Physio* reported between 5000 and 10,000 installs, only 86 users provided a rating (0.86-1.7%). No ratings, or less than the required minimum of 2 ratings, were provided for 30% (3/10) of the apps. All of the identified apps were designed exclusively for Android devices with no interoperability for iOS and Windows devices. This finding is in contrast to the 2015 systematic review of generic pain apps, where the majority of identified apps were designed for the iOS platform [15].

The adequacy of existing apps can be assessed relative to recently published clinical practice guidelines for evidence-based management of postoperative pain across the life span [23]. Relevant recommendations include the delivery of individually tailored preoperative education and perioperative pain management planning, the use of validated pain assessment tools to track response to pain interventions and inform treatment adjustments as needed, as well as the use of evidence-based, nonpharmacological pain management (eg, cognitive behavioral strategies, physical modalities) in conjunction with indicated pharmacological modalities [23]. Compared against these guidelines, none of the reviewed apps uniquely fulfill all recommended criteria of education (surgery and pain), symptom self-management, and pain self-management strategies. Furthermore, none of the apps offer the ability to prespecify a postoperative pain management plan or to consider the relevance of managing presurgical pain to postsurgical outcomes.

A complex myriad of surgical, psychological, socioenvironmental, and patient-related risk factors have been shown to influence postsurgical pain experience [24]. Pre- and postsurgical psychological factors associated with increased pain include anxiety, depression, low self-efficacy, and the tendency to catastrophize about pain [24-26]. Recent reviews and meta-analyses identify psychological treatments delivered preoperatively, postoperatively, or both, to effectively reduce postoperative pain [23,27]. Many of these strategies, such as distraction, relaxation, and guided imagery techniques, can be easily delivered via mHealth platforms. The ability of self-management apps to target psychosocial risk factors is plausible, given existing evidence supporting the use of smartphones to effectively deliver interventions to reduce symptoms of anxiety and depression [28], particularly when strategies are available in the moment when people are engaged in their everyday lives [29].

Despite the focus of many apps on physical strategies in recovery from surgery, none included goal setting. Goal setting

is a critical function in smartphone apps targeting increased physical activity [30]. In addition to increasing risk for postsurgical pain, depression, anxiety, and low self-efficacy are also identified as barriers to adherence to physiotherapy treatment [31]. Apps that are designed to pair psychological and physical pain self-management strategies with goal setting may be particularly effective for reducing pain and enhancing postsurgical outcomes.

Concurring with the findings of earlier app reviews [13,15,32], this study demonstrates that existing postoperative pain apps lack empirical evidence of evaluation. This finding continues to raise concerns about the trustworthiness and effectiveness of commercially available apps for helping individuals to manage their postoperative pain. In contrast to previous reviews, 50% (5/10) of the identified postoperative pain apps reported inclusion of a health care provider in their development. While this is a positive development, it would be informative for app developers to more explicitly describe the nature of provider input.

Moreover, none of the apps explicitly involved end users (people with postoperative pain) in their development. It has been recommended that patient-focused apps be developed as per a user-centered design (UCD) approach [33]. UCD is "characterized by a focus on the user, and on incorporating the user's perspective in all stages of the design process" [34]. Application of UCD has been associated with improved user acceptance, satisfaction, and engagement [35]. Given the limitations of existing apps, it is important that future app development to support postoperative pain management adopt the UCD principles.

Limitations

Given that the Web-based app searches were conducted within Canada, the results reflect the Canadian app marketplace. Due to the geographic search restrictions of the online app stores, this limitation also applies to other published app reviews (eg, Rosser and Eccleston [13] and Reynoldson et al [14] focused on the United Kingdom; Lalloo et al [15] focused on Canada). Thus, there may be differences in the commercially available postoperative pain apps in other countries. Additionally, while the Google Play Store (which contained all included apps) allows users to rate their app experience, their 1 to 5 scale does not have defined anchors. This lack of definition limits our ability to interpret the factors that are driving customer ratings. Furthermore, since app customers are not required to provide ratings, there is a high frequency of missing rating data in comparison with the reported number of installs.

Conclusions

Currently available postoperative pain apps are characterized by a lack of evidence-based content, goal setting, and social support functions and are not targeted to pediatric users. There is a need to develop and evaluate comprehensive, theory-based apps to better support patients with pain self-management care following surgery.

Conflicts of Interest

Authors CL, KAB, JNS, and FC conduct mHealth research and are involved in the development of pain self-management apps. These apps are described at the websites of PainQuILT (an evidence-based pain assessment and tracking tool for people with chronic pain) and the Improving Outcomes in Child Health through Technology (iOUCH) lab, a part of the Child Health Evaluative Sciences (CHES) research program at The Hospital for Sick Children in Toronto, Ontario.

References

1. Apfelbaum JL, Chen C, Mehta SS, Gan TJ. Postoperative pain experience: results from a national survey suggest postoperative pain continues to be undermanaged. Anesth Analg 2003 Aug;97(2):534-540. [Medline: 12873949]
2. Gan TJ, Habib AS, Miller TE, White W, Apfelbaum JL. Incidence, patient satisfaction, and perceptions of post-surgical pain: results from a US national survey. Curr Med Res Opin 2014 Jan;30(1):149-160. [doi: 10.1185/03007995.2013.860019] [Medline: 24237004]
3. Taylor RS, Ullrich K, Regan S, Broussard C, Schwenkglenks M, Taylor RJ, PAIN-OUT investigators. The impact of early postoperative pain on health-related quality of life. Pain Pract 2013 Sep;13(7):515-523. [doi: 10.1111/papr.12026] [Medline: 23279591]
4. MacLaren Chorney J, Twycross A, Mifflin K, Archibald K. Can we improve parents' management of their children's postoperative pain at home? Pain Res Manag 2014;19(4):e115-e123 [FREE Full text] [Medline: 25106030]
5. Pagé MG, Stinson J, Campbell F, Isaac L, Katz J. Identification of pain-related psychological risk factors for the development and maintenance of pediatric chronic postsurgical pain. J Pain Res 2013;6:167-180 [FREE Full text] [doi: 10.2147/JPR.S40846] [Medline: 23503375]
6. Connelly M, Fulmer RD, Prohaska J, Anson L, Dryer L, Thomas V, et al. Predictors of postoperative pain trajectories in adolescent idiopathic scoliosis. Spine (Phila Pa 1976) 2014 Feb 1;39(3):E174-E181. [doi: 10.1097/BRS.0000000000000099] [Medline: 24173016]
7. Parsons B, Schaefer C, Mann R, Sadosky A, Daniel S, Nalamachu S, et al. Economic and humanistic burden of post-trauma and post-surgical neuropathic pain among adults in the United States. J Pain Res 2013 Jun 17;6:459-469 [FREE Full text] [doi: 10.2147/JPR.S44939] [Medline: 23825931]

8. Madden M, Lenhart A, Duggan M, Cortesi S, Gasser U. Pewinternet. 2013. Teens and technology 2013 URL: http://www.pewinternet.org/files/old-media/Files/Reports/2013/PIP_TeensandTechnology2013.pdf [accessed 2017-06-02] [WebCite Cache ID 6qvDRSkXD]
9. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too!Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 6;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
10. 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]
11. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Construct validity and reliability of a real-time multidimensional smartphone app to assess pain in children and adolescents with cancer. Pain 2015 Dec;156(12):2607-2615. [doi: 10.1097/j.pain.0000000000000385] [Medline: 26580680]
12. Kirwan M, Vandelanotte C, Fenning A, Duncan MJ. Diabetes self-management smartphone application for adults with type 1 diabetes: randomized controlled trial. J Med Internet Res 2013 Nov 13;15(11):e235 [FREE Full text] [doi: 10.2196/jmir.2588] [Medline: 24225149]
13. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
14. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909 [FREE Full text] [doi: 10.1111/pme.12327] [Medline: 24422990]
15. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a Pain App for That": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
16. McGillion MH, Watt-Watson J, Stevens B, Lefort SM, Coyte P, Graham A. Randomized controlled trial of a psychoeducation program for the self-management of chronic cardiac pain. J Pain Symptom Manage 2008 Aug;36(2):126-140. [doi: 10.1016/j.jpainsymman.2007.09.015] [Medline: 18395397]
17. Stinson JN, McGrath PJ, Hodnett ED, Feldman BM, Duffy CM, Huber AM, et al. An internet-based self-management program with telephone support for adolescents with arthritis: a pilot randomized controlled trial. J Rheumatol 2010 Sep;37(9):1944-1952. [doi: 10.3899/jrheum.091327] [Medline: 20595280]
18. Palermo TM, Eccleston C, Lewandowski AS, Williams AC, Morley S. Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review. Pain 2010 Mar;148(3):387-397 [FREE Full text] [doi: 10.1016/j.pain.2009.10.004] [Medline: 19910118]
19. Lorig K, Holman H. Arthritis self-management studies: a twelve-year review. Health Educ Q 1993;20(1):17-28. [Medline: 8444622]
20. Ontario. Regulated Health Professions Act, 1991 URL: https://www.ontario.ca/laws/statute/91r18 [accessed 2017-06-02] [WebCite Cache ID 6qvFMqvmw]
21. de la Vega R, Miró J. mHealth: a strategic field without a solid scientific soul. a systematic review of pain-related apps. PLoS One 2014 Jul;9(7):e101312 [FREE Full text] [doi: 10.1371/journal.pone.0101312] [Medline: 24999983]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]
23. Chou R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, et al. Management of postoperative pain: a clinical practice guideline from the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain 2016 Feb;17(2):131-157. [doi: 10.1016/j.jpain.2015.12.008] [Medline: 26827847]
24. Katz J, Seltzer Z. Transition from acute to chronic postsurgical pain: risk factors and protective factors. Expert Rev Neurother 2009 May;9(5):723-744. [doi: 10.1586/ern.09.20] [Medline: 19402781]
25. Rabbitts JA, Fisher E, Rosenbloom BN, Palermo TM. Prevalence and predictors of chronic postsurgical pain in children: a systematic review and meta-analysis. J Pain 2017 Jun;18(6):605-614. [doi: 10.1016/j.jpain.2017.03.007] [Medline: 28363861]
26. Theunissen M, Peters ML, Bruce J, Gramke HF, Marcus MA. Preoperative anxiety and catastrophizing: a systematic review and meta-analysis of the association with chronic postsurgical pain. Clin J Pain 2012;28(9):819-841. [doi: 10.1097/AJP.0b013e31824549d6] [Medline: 22760489]
27. Davidson F, Snow S, Hayden JA, Chorney J. Psychological interventions in managing postoperative pain in children: a systematic review. Pain 2016;157(9):1872-1886. [doi: 10.1097/j.pain.0000000000000636] [Medline: 27355184]
28. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
29. Loo GB, Griffiths KM, Gulliver A. Effectiveness of mobile technologies delivering ecological momentary interventions for stress and anxiety: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):221-229. [doi: 10.1093/jamia/ocv043] [Medline: 25997643]

30. 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 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
31. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228 [FREE Full text] [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
32. Wallace LS. A review of pain-related smartphone apps. 2013 Presented at: American Academy of Pain Medicine Annual Meeting; 2013; Fort Lauderdale, FL. [doi: 10.1111/pme.12104]
33. Steele Gray C, Khan AI, Kuluski K, McKillop I, Sharpe S, Bierman AS, et al. Improving patient experience and primary care quality for patients with complex chronic disease using the electronic patient-reported outcomes tool: adopting qualitative methods into a user-centered design approach. JMIR Res Protoc 2016 Feb 18;5(1):e28 [FREE Full text] [doi: 10.2196/resprot.5204] [Medline: 26892952]
34. Devi KR, Sen A, Hemachandran K. A working framework for the user-centered design approach and a survey of the available methods. Int J Sci Res Publ 2012 Apr;2(4):1-8 [FREE Full text]
35. Chan J, Shojania KG, Easty AC, Etchells EE. Does user-centred design affect the efficiency, usability and safety of CPOE order sets? J Am Med Inform Assoc 2011 May 1;18(3):276-281 [FREE Full text] [doi: 10.1136/amiajnl-2010-000026] [Medline: 21486886]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UCD: user-centered design

Edited by G Eysenbach; submitted 22.06.17; peer-reviewed by J Prescott, M Mittinty; comments to author 13.07.17; revised version received 14.09.17; accepted 20.09.17; published 23.10.17.

Please cite as:
Lalloo C, Shah U, Birnie KA, Davies-Chalmers C, Rivera J, Stinson J, Campbell F
Commercially Available Smartphone Apps to Support Postoperative Pain Self-Management: Scoping Review
JMIR Mhealth Uhealth 2017;5(10):e162
URL: http://mhealth.jmir.org/2017/10/e162/
doi:10.2196/mhealth.8230
PMID:29061558

Original Paper

Posttraumatic Stress Disorder and Mobile Health: App Investigation and Scoping Literature Review

Carolina Rodriguez-Paras[1*], MS; Kathryn Tippey[2*], PhD; Elaine Brown[3*], MA; Farzan Sasangohar[1,4,5*], PhD; Suzannah Creech[6*], PhD; Hye-Chung Kum[1,4,7*], PhD; Mark Lawley[1,4*], PhD; Justin K Benzer[6,8*], PhD

[1]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[2]Center for Research and Innovation in Systems Safety, Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States

[3]Health Science Center, School of Public Health, Louisiana State University, New Orleans, LA, United States

[4]Center for Remote Health Technologies and Systems, Texas A&M University, College Station, TX, United States

[5]Department of Environmental and Occupational Health, School of Public Health, Texas A&M University, College Station, TX, United States

[6]VISN 17 Center of Excellence for Research on Returning War Veterans, Central Texas Veterans Health Care System, Waco, TX, United States

[7]Department of Health Policy and Management, School of Public Health, Texas A&M University, College Station, TX, United States

[8]Department of Psychiatry, Dell Medical School, University of Texas, Austin, TX, United States

[*]all authors contributed equally

Corresponding Author:
Farzan Sasangohar, PhD
Department of Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX, 77843
United States
Phone: 1 9794582337
Email: sasangohar@tamu.edu

Abstract

Background: Posttraumatic stress disorder (PTSD) is a prevalent mental health issue among veterans. Access to PTSD treatment is influenced by geographic (ie, travel distance to facilities), temporal (ie, time delay between services), financial (ie, eligibility and cost of services), and cultural (ie, social stigma) barriers.

Objective: The emergence of mobile health (mHealth) apps has the potential to bridge many of these access gaps by providing remote resources and monitoring that can offer discrete assistance to trauma survivors with PTSD and enhance patient-clinician relationships. In this study, we investigate the current mHealth capabilities relevant to PTSD.

Methods: This study consists of two parts: (1) a review of publicly available PTSD apps designed to determine the availability of PTSD apps, which includes more detailed information about three dominant apps and (2) a scoping literature review performed using a systematic method to determine app usage and efforts toward validation of such mHealth apps. App usage relates to how the end users (eg, clinicians and patients) are interacting with the app, whereas validation is testing performed to ensure the app's purpose and specifications are met.

Results: The results suggest that though numerous apps have been developed to aid in the diagnosis and treatment of PTSD symptoms, few apps were designed to be integrated with clinical PTSD treatment, and minimal efforts have been made toward enhancing the usability and validation of PTSD apps.

Conclusions: These findings expose the need for studies relating to the human factors evaluation of such tools, with the ultimate goal of increasing access to treatment and widening the app adoption rate for patients with PTSD.

(JMIR Mhealth Uhealth 2017;5(10):e156) doi:10.2196/mhealth.7318

KEYWORDS

posttraumatic stress disorders; PTSD; mobile health; mHealth; anxietys

Introduction

Mobile Health (mHealth) Apps

Recent technological advances have resulted in the development of emerging mobile health (mHealth) apps. mHealth apps include a wide range of applications such as educational materials and self-management platforms, health care–specific tools for managing the therapeutic process, health and preventative behavior, patient and patient-provider roles and relationships, challenges of daily life, and crisis situations. In addition, many of these apps are free and can help to reduce barriers to access, as they are able to provide the assistance to patients at any given time, whereas speaking with a clinician requires an appointment, traveling to the facility, and financial considerations. mHealth apps have thus far been successfully implemented across a wide range of medical disciplines, including dermatology [1], ophthalmology [2], and nutritional sciences [3] and have addressed many specific illnesses such as diabetes [4] and infectious diseases [5]. Apps have also been developed to aid in the treatment of mental disorders such as posttraumatic stress disorder (PTSD) [6-10].

Posttraumatic Stress Disorder

PTSD is a prevalent mental health issue that commonly occurs after a person has experienced a traumatic event, which can include being threatened with death or experiencing the death of others (eg, death of a family member or a friend), sexual violence, or serious injury [11]. Approximately, 7.8% of the American population will experience PTSD at some point during their lifetime [12], with veterans being between 5% and 25% more likely to experience PTSD depending on the service era in which they served their country (ie, Vietnam War, Gulf War, Operations Iraqi Freedom, Enduring Freedom, and New Dawn) [11,13]. All diagnoses of PTSD must occur after a traumatic experience, but not all traumatic experiences lead to the development of PTSD. Diagnosis of PTSD involves symptoms from each of the following four clusters: (1) intrusion (eg, nightmares and flashbacks), (2) avoidance (eg, thoughts and feelings), (3) negative alterations to cognitions and mood, and (4) alterations to arousal and reactivity (eg, depression or sleep deprivation) [14]. As noted, military veterans are a special population whose warzone experiences may increase the risk for PTSD. Furthermore, many veterans diagnosed with PTSD still serve the military in active combat roles, with this number peaking at over 17,000 servicemen in 2012 [15]. However, evidence has supported the efficacy of multiple PTSD treatment options in reducing overall posttraumatic stress symptoms in veterans diagnosed with PTSD [16-18].

PTSD Treatment

Current PTSD treatments are divided into two categories that are not mutually exclusive: (1) pharmacotherapy and (2) psychotherapy. Studies suggest psychotherapy is more effective than pharmacotherapy [16], with two cognitive behavioral therapy (CBT) treatment methods considered the most effective: (1) prolonged exposure (PE) therapy and (2) cognitive processing therapy (CPT) [16]. PE was developed from the idea of *prolonged* or *repeated* exposure to traumatic events [17]. Patients repeatedly encounter situations known to cause symptoms while in a safe environment, with the expectation that they will overcome their fears [17]. CPT helps patients understand and change the way they think about traumatic events by emphasizing that they are not at fault [18].

Despite the efficacy of current PTSD therapies, significant challenges still exist in access to treatment, particularly for veterans, that may influence both treatment-seeking behavior and adherence to treatment [19,20]. First, access to mental health care is becoming increasingly more difficult. Mental health facilities continue to be understaffed despite the actively growing demand for mental health care [21], with reports indicating that facilities serving veterans have an insufficient mental health workforce to meet the needs of personnel returning home from active duty [21,22]. Access to health care may be influenced by geographic, temporal, financial, cultural, and technological factors [23]. Geographic factors create difficulties in traveling to mental health facilities. Temporal factors include obstacles in scheduling care (eg, evening or weekend appointments). Financial factors center on the limitations of health insurance, co-pays, and other monetary resources needed to attend clinical sessions. Cultural factors stem from the societal stigma associated with receiving mental health care [24-26] and may be a particularly important factor in dissuading veterans from seeking care because of the military culture related to mental health care [27]. Finally, technological factors involve the different barriers in obtaining or using technologies.

mHealth apps have the potential to help improve access to care [28], foregoing some of the difficulties in accessing mental health care as stand-alone tools (the patient can use the app without a clinician) or tools used in coordination with the clinician or in conjunction with the clinical treatment. Apps may effectively address technological access challenges as smartphones are widely available (64% of Americans owned a smartphone in 2015) [29], and apps can make it easier for patients to access information about the diagnosis and treatment of mental disorders. Apps may address cultural access by offering patients a discreet mobile environment to manage their disorder. In addition, apps may lessen geographic access issues by decreasing the number of in-person appointments needed. A particular strength of apps is the potential to improve temporal access by allowing patients and providers to conduct some of their therapeutic work asynchronously. Patients can use apps to monitor and manage their symptoms, record and replay therapy sessions, connect with clinicians or emergency personnel in the case of crises [30], and engage in social connections with communities of trauma survivors for additional treatment support [31]. Clinicians may use apps to collect data on patient engagement and progress during and between sessions, such as how much time the patient spends doing practice assignments, how often patients utilize emotion regulation skills such as relaxation, and how well the treatment is helping to manage their symptoms. Along these same lines, the addition of external wearable sensors, such as those found in smartwatches, may provide both patients and clinicians with additional data on sleep quality and could potentially aid in determining events that trigger hyperarousal [7].

The goal of PTSD apps should be to aid in the treatment and monitoring of trauma survivors with PTSD and to provide both

patients and clinicians with timely remote feedback that can supplement or enhance current therapies [10]. The potential benefits of properly validated PTSD apps are to effectively engage trauma survivors with PTSD, thus improving their access to care [32]. However, limited knowledge exists about the availability of apps for PTSD and about how well these apps were developed. The aim of this paper was to document the currently available PTSD-related apps as well as the usage of these apps and validation procedures used when designing them.

Methods

Overview of Research Methods

This study consists of two parts: (1) a review of publically available PTSD apps, which includes more detailed information about the three most prevalent apps used and (2) a scoping literature review performed using systematic methods. The purpose of part 1 is to determine the availability of PTSD apps, and the purpose of part 2 is to determine the usage and efforts toward validation of such mHealth apps.

mHealth App Search Method

Health care providers' mobile app websites (eg, the Department of Veterans Affairs [VA] App Store), commercial app stores (eg, Apple App Store, Google Play Store), websites that aggregated or listed mental health apps, websites that provide app ratings, Web communities supporting veterans, and Google were used to search keywords relating to [post-traumatic stress disorder" OR "PTSD"], ["veterans"], and words relating to PTSD treatment (eg, ["insomnia"]). The search was performed from January 2016 to August 2016. Data on available apps were collected from the following websites: Apple App Store [33], Google Play Store [34], VA App Store [35], National Center for Telehealth & Technology mobile applications site [36], and Amazon App store for Android [37]. All apps were available on at least one of these websites as of August 10, 2016. The following inclusion criteria was used: the app must be relevant to PTSD, PTSD treatment, or common symptoms of PTSD (ie, depression, anxiety, insomnia, and anger); treatment apps related to veteran support must be specific to the US military or US veterans; and the app had to be able to be used without opening a Web browser. Apps may include content related to mental health disorders that are comorbid with PTSD. For example, apps may address both PTSD and depression. However, no apps that solely addressed depression, and not PTSD, were included.

Apps were categorized according to clinical focus (ie, mental health disorder, PTSD symptom, or clinical treatment modality) and for each type of app utility used (eg, education and exercises). Each app was assigned to one primary clinical focus but could contain more than one app utility. For example, Acceptance and Commitment Therapy (ACT) Coach included mindfulness exercises, but the primary focus of this app is specific to a VA-recognized evidence-based treatment for PTSD. Apps may use several utilities. For example, PE Coach included multiple utilities such as educational materials and exercises. Apps within each clinical focus and app utility categories were tallied to determine frequency. The following information for each app was also collected to determine feasibility and acceptability: average user ratings, number of user ratings, availability to iPhone operating system (henceforth iOS) (Apple, Cupertino, CA) and/or Android (Google, Mountain View, CA) operating systems (henceforth Android), minimum iOS and/or Android requirements, and cost to download.

Scoping Literature Review Method

The app search did not reveal information on how the apps were designed and evaluated, or whether studies had analyzed their usability. A scoping literature review [38] was conducted to further understand the scope of available evidence and support the efficacy of such tools.

A combination of keywords relating to ["post-traumatic stress disorder" OR "PTSD"] AND ["mobile applications" OR "mHealth"] was used to search within Google Scholar and the Texas A&M EBSCOHost Research Databases such as MEDLINE, ABI/INFORM Complete, and Academic Search Complete. The Google Scholar database search was completed on March 16, 2016, with a total of 1850 results. The search using other databases did not result in any new results. The search only included journals written in English and published in or after 2011, as this is the year in which PTSD Coach, the first of the VA PTSD mobile apps, first appeared on the app store. The following inclusion criteria were then used to narrow the scope of the papers obtained through the search: the paper reviewed or validated an existing PTSD app (eg, feasibility studies, randomized clinical trials, usability testing, etc), the paper detailed the development of a new app for the detection or treatment of PTSD, or the paper was a case study using PTSD apps.

Results

The following section first details the results from the review of publically available PTSD apps and then the scoping literature review.

mHealth App Search Results

A total of 201 apps were chosen for the study, all of which were available for iOS or Android. Apps were categorized based on their primary focus or purpose, which fell into six distinct groups. These were as follows: (1) PTSD evidence-based treatment (EBT), which included apps specific to a VA-recognized EBT for PTSD [18]; (2) PTSD-specific, which included apps that provided educational materials, exercises, and/or symptom tracking only for PTSD but were not specific to any evidence-based PTSD treatment method utilized at VA; (3) general mental health (MH) that included PTSD, which included educational materials and exercise content on multiple mental health conditions (eg, depression and anxiety), as well as content related to PTSD; (4) mindfulness and relaxation techniques commonly used with PTSD, which included content specific to mindfulness techniques or exercises; although these apps are not specific to PTSD or a treatment for PTSD, mindfulness techniques are a core component of ACT therapy, which is recognized at VA as an EBT for PTSD; (5) anger management, which is a common skills deficit in patients who have PTSD, and therefore, commonly addressed in treatment for PTSD; these apps were specific to anger management but were not specific to PTSD; (6) insomnia, another common

symptom of PTSD, is also a frequent focus in treatment for PTSD; these apps were specific to addressing insomnia but were not specific to PTSD (Table 1).

The total number of 201 apps chosen included duplicate apps between operating systems; for example, PTSD Coach was available for both iOS and Android, so it was counted twice. When duplicates were removed, the total number of apps was 81. Across all categories, apps related to mindfulness and relaxation were the most frequently available (approximately 29.9%, 60/201) followed by PTSD-specific apps (approximately 22.4%, 45/201).

Apps were further analyzed for content and utilities, which included the following: (1) educational information, which included educational material regarding PTSD, PTSD treatment, or common symptoms of PTSD; (2) exercises, which included either skills training or practice components; (3) symptom tracking, which included tools to track severity of symptoms of PTSD, whether or not the app was related to PTSD (eg, tracking sleep); (4) connections to outside professional support, which provided methods of contacting outside professional support, and this included direct contact options (eg, send a message through the app) or providing contact information (eg, phone number); (5) connections to outside peer support, which included content for individuals with PTSD for communication with local or online peer support; and (6) components specific to treatment integration, which included content designed to be integrated into ongoing in-person treatment with a therapist. An example of treatment integration content included patients' ability to audio-record in-person therapy sessions and replay them outside of therapy as part of a treatment homework assignment (Table 2). Tallies of app content categories that are included in Table 2 do not include duplicates between operating systems. For example, CPT Coach utilities were only tallied once and not twice for CPT Coach for iOS and for Android.

Across utilities, in-app exercises (eg, guided breathing) were the most commonly provided app function and were especially dominant across apps related to mindfulness and relaxation. All apps that were directly related to an EBT for PTSD included exercise components, and apps specific to PTSD commonly utilized both education and exercise components. Both PTSD-specific and mindfulness and relaxation app categories had at least one app to provide for one or more of all measured utility categories. All app categories included at least one app that provided education.

Building on this work, app accessibility was measured in two ways: (1) technological access, defined as the minimum operating system and memory required to use the app, with lower operating systems and less memory requirements having better accessibility and (2) financial access, defined as the cost of downloading the app, with lower costs having better accessibility. With respect to *technological access*, although some Android apps adjusted to the operating system on the user's phone, several Android and iOS apps required relatively recent versions of operating systems (eg, iOS 8.1), which may act as a barrier to individuals with older smartphones with low storage space.

Table 1. The app categories grouped by the type of operating system.

Type of operating system	PTSD[a] EBT[b]	PTSD-specific	MH variety	Mindful and relax	Anger	Insomnia	Total
iPhone operating system (iOS) only	8	12	8	22	4	12	66
Android only	7	14	5	16	4	8	54
Both iOS and Android	12	19	8	22	8	12	81
Total	27	45	21	60	16	32	201

[a]Posttraumatic stress disorder.

[b]Evidence-based treatment.

[c]MH: mental health.

Table 2. The app tallies for different utility categories (utilities are not mutually exclusive).

App categories (total from search)	Education	Exercises	Tracking	Professional support	Peer support	Treatment integration	Other
PTSD[a] EBT[b] (27)	5	12	4	2	0	2	2
PTSD-specific (45)	13	12	5	4	2	0	9
MH variety (21)	2	5	5	0	1	0	8
Mindful and relax (60)	7	22	7	2	3	0	19
Anger (16)	5	7	2	0	0	0	2
Insomnia (32)	3	11	3	0	1	0	14
Total	35	69	26	8	7	2	54

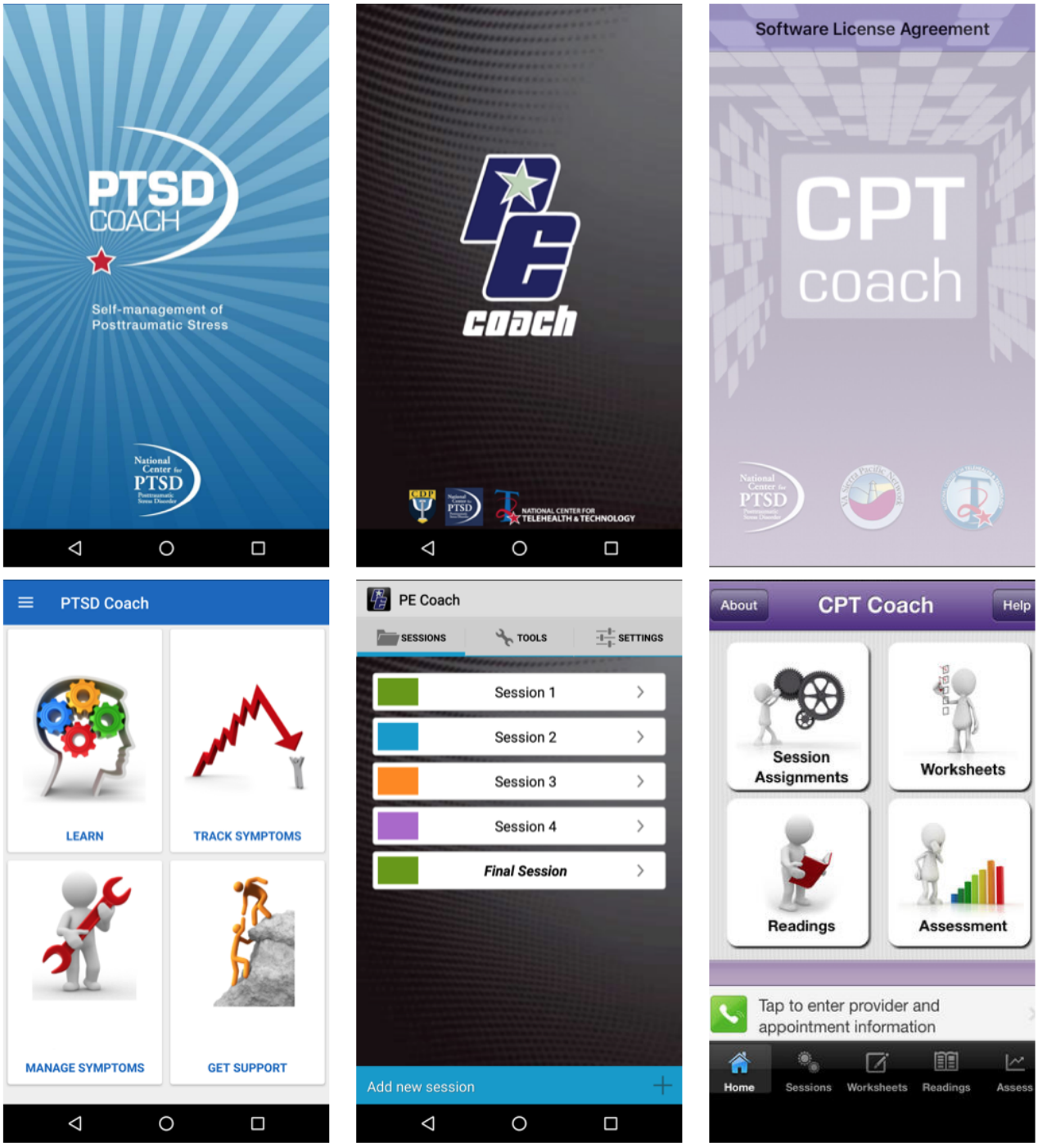
[a]Posttraumatic stress disorder.

[b]Evidence-based treatment.

For example, an Apple phone with 8 GB of space may not be able to download an upgrade that requires over 4 GB of available space. Regarding *financial access*, in 2013, a new smartphone cost, on average, US $531 in North America (Can $662.13, Aus $678.28) [39], and, for most major smartphone carriers, the first 1 to 3 GB of data costs an average of US $35 per month (Can $43.64, Aus $44.71) [40,41]. Most apps were either free or inexpensive to download, with the most expensive apps at US $9.99 (Can $12.46 , Aus $44.71). However, several apps included in-app purchases or equipment that cost much more than the app itself.

Three mHealth apps reviewed in this search appeared dominant among users (most downloads), all of which were developed by the Department of Veterans Affairs for trauma survivors with PTSD: PTSD Coach (261,045 total downloads), PE Coach (49,453 total downloads), and CPT Coach (11,689 total downloads). Download counts were reported by J Worthen from the National Center for Telehealth and Technology (September 12, 2016). These apps can be divided into two categories based on their intended use: (1) as a stand-alone app for the self-management of symptoms (ie, PTSD Coach) or (2) in conjunction with a PTSD EBT through a health care provider (ie, PE Coach and CPT Coach). PE Coach and CPT Coach are the only apps designed thus far explicitly for integration with standard treatment.

Figure 1. PTSD Coach (left), PE Coach (middle), CPT Coach (right).



PTSD Coach (Figure 1, left) is an mHealth app designed to explain PTSD concepts to patients and provide them with self-management tools based on CBT. Patients can use the tools in this app to learn, perform self-assessments, manage symptoms, and find support. The *learn* section provides trauma survivors and their family members with information about PTSD. The app's *self-assessment* and *manage symptoms* sections provide individuals with a checklist to assess the severity of their symptoms, which patients can track over time to see their treatment progress and a list of mitigation techniques to cope with distressing situations, such as prompts to think about pleasant events and guidance through progressive muscle relaxation. After users finish using each mitigation technique, they are prompted to complete the checklist again; if the person rates their distress the same or higher, then they are offered another mitigation tool to try. The *find support* section allows trauma survivors to store contact information for those they rely on during emergency or crisis situations, making this information easy to access.

Of the 201 apps collected, only 2 were designed explicitly for integration with standard treatment for PTSD: PE Coach (with PE therapy; Figure 1, middle) and CPT Coach (with CPT; Figure 1, right). PE Coach is divided according to the different sessions for PE. For example, Session 1 contains only the PTSD Checklist assessment, the option to record the session, scheduling the next appointment, and the assigned homework, whereas Session 2 has the additional options to review previous homework assignments, add anchors for Subjective Units of Distress Scale, in vivo hierarchy, and in vivo homework, which occur only after the first session is complete. After Session 3, PE Coach also allows patients to record the imaginal exposure portion of the session, where the person processes the memory which causes their PTSD symptoms. Similarly, CPT Coach has information detailing the treatment components of CPT, homework assignments, and exercises. Both these apps also offer benefits to clinicians by allowing them to review a patient's homework or monitor how much the patient is using the app during the session, allowing providers to see the patient's between-session progress and to review these data with the patient during the session. In addition, the apps provide patients with reminders or help them schedule follow-up appointments.

Scoping Literature Review Results

A total of 1850 papers were found. The 28 papers fitting the inclusion criteria are listed in Table 3. Table 3 also indicates whether or not each paper addresses some relevant usage categories identified in this research. The categories were defined to understand the human factors approach to the app design and development, including the population, any analysis performed on the app, the user demographics, and the usage and adoption. The categories were not mutually exclusive; thus, a paper could fit into more than one designation.

The first category "Veteran population" refers to papers that specifically mention PTSD apps in reference to the veteran populations (as opposed to the general population). "Benefit analysis" contains the papers that focus on the perception or evaluation of app, or with the potential benefits of apps to the population. "Age concern" includes the papers that discuss potential barriers to older populations, or the appeal of mHealth apps to the younger populations. "Usage and adoption" refers to papers that specifically mention how the users have interacted with the apps, particularly those that contain analytics on the use of apps, such as statistics about downloads, how often users return to the app after the initial download, and the number of users who download the app specifically as part of their treatment for PTSD. The final category "HFE considerations" (ie, human factors and ergonomics) contains the papers that mention app design concerns, including usability of apps, user satisfaction, and acceptability of apps, or other HFE analysis beyond that in the previous categories.

Although all the papers listed in Table 3 met the inclusion criteria, the majority only mentioned PTSD apps as an example or in passing. The 6 papers with the footnote in the reference column in Table 3 focused on more detailed analysis of a specific PTSD app—either PE Coach or PTSD Coach—discussing their usage or validation more extensively than the other papers and hence are discussed here in further detail.

Three studies examined PE Coach. Reger et al [44,60] and Kuhn et al [59] explored the functions and potential benefits of using PE Coach as an adjunct to traditional therapy. Studies conducted before the app's release highlighted the features of the app [44] and surveyed clinicians' perceptions on the usefulness of the app [59]. The only study performed after PE Coach's release was Reger et al [60], which was a case study examining two soldiers' perceptions of the usability and their satisfaction with the app when implemented along with PE therapy. Although all these studies had positive results, most noted the need for additional testing to determine the app's impact on clinical outcomes.

Three studies examined PTSD Coach, all of which were more analytical than those that examined PE Coach. Both Kuhn et al [9] and Owen et al [42] explored app usage, with Kuhn et al [9] focusing more on subjective perceptions, such as user satisfaction and helpfulness, and Owen et al [42] focusing more on objective user engagement via Flurry Analytics software (Yahoo! Developer Network). Owen et al [42] was also able to determine common points of attrition in using the app. The final paper [46] was a randomized control trial evaluating the effectiveness of PTSD Coach as a stand-alone app versus when used in conjunction with clinical support; both treatment groups had clinically significant improvements in their symptoms but having clinician support resulted in greater reductions in PTSD symptoms.

Table 3. The 28 papers meeting the inclusion criteria. The headers are the categories of gaps identified. An "X" in the column indicates that the papers worked to address this gap.

Paper number and author name		Veteran population	Benefit analysis	Age concern	Usage and adoption	HFE[a] considerations
1	Erbes et al [6]	X	X	X	X	X
2	Gravenhorst et al [7]			X		X
3	Chen et al [8]	X				
4	Kuhn et al [9][b]	X	X	X	X	X
5	Olff [10]					
6	Sloan et al [30]	X				
7	Owen et al [42][b]	X	X		X	X
8	Kuhn et al [43]		X	X		X
9	Reger et al [44][b]	X				
10	McInnes et al [45]	X		X		X
11	Possemato et al [46][b]	X	X			
12	Gratzer et al [47]					
13	Kuester et al [48]					
14	Luxton et al [49]					
15	Fletcher et al [50]		X			
16	Castro et al [51]	X				
17	Turvey et al [52]					
18	Kanuri et al [53]					X
19	Baysari et al [54]					
20	Mohsenin et al [55]					
21	Driesenga et al [56]	X				
22	Proudfoot [57]					
23	Price et al [58]	X				
24	Kuhn et al [59][b]		X	X		X
25	Reger et al [60][b]	X	X			X
26	Olff et al [61]		X			X
27	Weingardt and Greene [62]	X				
28	Chan et al [32]					

[a]Human factors and ergonomics.

[b]Papers that provide more detailed analysis of a specific PTSD App (PE Coach or PTSD Coach).

To date, no papers exist that compare the stand-alone app (PTSD Coach) with the adjunctive apps (CPT Coach and PE Coach). Although PTSD Coach offers more information on PTSD and some therapy tools, both PE and CPT Coach contain information relevant for each clinical session. For example, these adjunctive apps may be able to improve treatment adherence by offering reminders on homework assignments due. Clinicians' perceptions tend to be favorable toward implementing apps for PTSD treatment, such as PE Coach [43]. Similar apps that address mental or chronic conditions include those for helping users manage stress [63] and for aiding arthritis patients in medication and exercise compliance [64].

Some of this literature conceptualizes mHealth apps as a method for overcoming geographic and temporal access barriers to mental health care [42]. Although there are few papers evaluating the PTSD apps' effectiveness, the papers reviewed indicated that mobile phones may be an adequate supplement for PTSD treatment [7]. Increasing temporal access through mHealth apps to health services may also result in lower cost, improved patient satisfaction, and improved health outcomes [8].

Discussion

Principal Findings

The goal of this review was to determine the availability and level of validation of PTSD apps. The mHealth app search indicated that a plethora of Android and iOS PTSD-specific apps are available. In addition, the app search showed that other PTSD-related apps such as relaxation, insomnia management, and anger management tools are available but may not be necessarily known to PTSD patients. Despite the availability of these tools, the scoping literature review suggested there was insufficient evidence on the validity of the apps. The literature review highlighted the need for additional studies on app dissemination and adoption [6,32,44,45,49,53] and app validation and treatment integration [7,43,44,46-48,52,59]. This literature also suggests the potential of integrating such PTSD-related apps with new technologies, such as smartwatches, as an adjunct for improving treatment.

Of the PTSD-related apps available, all of the VA's apps including the three most-downloaded apps for PTSD treatment (PTSD Coach, CPT Coach, and PE Coach) are free to download. However, the results suggest many trauma survivors with PTSD, and even some practitioners, may not be aware that these apps exist [8,6,10,52,57]. Many factors may influence app adoption, including social contacts [65]. In addition, there are different business models that can be implemented depending on the type of app [66]. Additional concerns over app usage remain, focusing on issues relating to patient data confidentiality, data storage, legal and ethical issues [67], and on issues relating to the cost of smartphones and data plans, app software requirements, and the cost of some apps developed by non-VA developers. More information is hence needed on the dissemination and adoption of these mHealth tools and how to improve their accessibility to PTSD patients.

This review found only 6 papers analyzing the usage or validation of specific PTSD-related apps [9,42,44,46,59,60]. Of the 6 papers examined in further detail, none provided evidence of the implementation of user-centered procedures during the app development to improve usability. High-level usability fosters user engagement and is essential to ensure such apps reach a broad audience of veterans that include those with PTSD receiving minimal or insufficient treatment [68]. Evidence also suggests the majority of the PTSD-related apps have yet to be validated, with no validation efforts reported for these apps in the patient-provider context before their release [44]. These papers also highlighted the small sample sizes in user studies conducted postapp release [9,31,60]. This lack of documentation of the user-centered design and testing process is less rigorous compared with traditional validation procedures set out by the US Food and Drug Administration. An important question to ask is what should be the level of validation needed for these apps. App validation will incur into additional time and monetary resources for app developers. However, the lack of validation data limits the ability of consumers to determine which apps may be helpful in managing PTSD symptoms. A good balance might be to conduct selective postrelease validation studies on popular apps.

The potential legal ramifications posed by mHealth apps center on scrutiny of data storage and remote communication with health care providers. The responsible party for these legal issues—whether it is the developer, the health care provider, the user, or some combination of the aforementioned—has yet to be determined [7]. Current PTSD-related apps reside solely on the smartphone and involve no data communication with any hospital system. The threat of legal ramifications for data security issues relating to remote communication is the primary reason why many PTSD-related apps do not have user accounts or remote connections with clinicians [8]. Data synchronization issues within complex health care environments and accountability after automatic data notifications contribute to health care providers' concerns.

Despite potential security issues and obstacles to apps' dispersion among users, an abundance of calls for more studies on app dissemination and adoption [6,32,44,45,47,49,53] and on app validation and treatment integration [7,43,44,46-48,52,59] indicates the importance of developing apps with optimal usability and improve treatment. Of those apps designed to be used in conjunction with clinical treatment, only two currently incorporate specific treatment-related components (PE Coach and CPT Coach). This indicates a need for further app development and refinement to include treatment integration components such as objective measures of PTSD. These objective measures may be provided through adding functions incorporating wearable sensors into PTSD-related apps, which would provide valuable information to aid in treatment.

Recent advances in wearable sensors provide an opportunity for future PTSD apps to utilize the features of activity trackers for tracking patients' PTSD-related physiological changes, thus aiding in diagnosing, monitoring, and optimizing clinical treatment regimens. This has the potential to significantly contribute to on-going therapies. App-integrated information from wearable sensors could be used to inform patients of their current physiology and mental state or provide clinicians with information about the most viable treatment course or aid in remote care [54]. Voice modulation has successfully been used to identify depression and anxiety, indicating its potential to detect PTSD triggers [69]. Physiological monitors, such as heart rate straps and watches, can also connect to mobile apps and act as an indicator of PTSD trigger [50,70], particularly because trauma survivors with PTSD tend to have a higher heart rate variability, which is an autonomic indicator of how they cope with stress that is currently being independently tested for biofeedback therapy [70].

The mHealth app search centered around four main symptoms of PTSD (eg, intrusion, avoidance, negative alterations to mood and cognition, and changes in arousal and reactivity), which are common among mental health issues including PTSD. Apps specific to other mental conditions that may aid in PTSD were hence not included in this analysis. In addition, although we focused on stand-alone mobile apps, there are several mobile-responsive websites related to PTSD that were not included in this search. Although accessibility was assessed and discussed based on technical and financial access, future studies should investigate design for the disabled (eg, blind, cognitively

impaired, deaf, and users with missing limbs). Although the literature review yielded only a small number of peer-reviewed publications relating to PTSD mobile apps, the combined results of the app search and review of literature shed light on the current state of mHealth apps to support PTSD patients.

Conclusions

This dual review highlights the availability and potential of PTSD app usage in increasing treatment adherence and quality. Results of this review suggest that current app development, however, lacks strong usability and validation components and that not enough apps are being developed to be integrated as treatment tools. These findings expose the need for studies relating to the human factors evaluation of such tools, with the ultimate goal of increasing access to treatment and widening the app adoption rate, for patients with PTSD.

Acknowledgments

This research was funded by a Texas A&M Engineering Experiment Station (TEES) Grant and a Department of Veterans Affairs Career Development Award (CDA 11-246). The views expressed in this paper are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US Government. The authors would like to thank the members of the Applied Cognitive Ergonomics Laboratory at Texas A&M University for their insightful feedback.

Conflicts of Interest

None declared.

References

1. Robson Y, Blackford S, Roberts D. Caution in melanoma risk analysis with smartphone application technology. Br J Dermatol 2012 Sep;167(3):703-704. [doi: 10.1111/j.1365-2133.2012.11046.x] [Medline: 22762381]
2. Bhosai SJ, Amza A, Beido N, Bailey RL, Keenan JD, Gaynor BD, et al. Application of smartphone cameras for detecting clinically active trachoma. Br J Ophthalmol 2012 Oct;96(10):1350-1351 [FREE Full text] [doi: 10.1136/bjophthalmol-2012-302050] [Medline: 22797320]
3. 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]
4. Ciemins E, Coon P, Sorli C. An analysis of data management tools for diabetes self-management: can smart phone technology keep up? J Diabetes Sci Technol 2010 Jul 01;4(4):958-960 [FREE Full text] [doi: 10.1177/193229681000400427] [Medline: 20663462]
5. Oehler RL, Smith K, Toney JF. Infectious diseases resources for the iPhone. Clin Infect Dis 2010 May 01;50(9):1268-1274. [doi: 10.1086/651602] [Medline: 20233061]
6. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014 Nov;179(11):1218-1222. [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
7. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. Pers Ubiquit Comput 2014 Sep 21;19(2):335-353. [doi: 10.1007/s00779-014-0829-5]
8. Chen C, Haddad D, Selsky J, Hoffman JE, Kravitz RL, Estrin DE, et al. Making sense of mobile health data: an open architecture to improve individual- and population-level health. J Med Internet Res 2012 Aug 9;14(4):e112 [FREE Full text] [doi: 10.2196/jmir.2152] [Medline: 22875563]
9. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
10. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015 May 19;6:27882 [FREE Full text] [Medline: 25994025]
11. Ptsd.va. How common is PTSD? URL: http://www.ptsd.va.gov/public/PTSD-overview/basics/how-common-is-ptsd.asp [accessed 2017-01-12] [WebCite Cache ID 6nT31yFoo]
12. Goldstein RB, Smith SM, Chou SP, Saha TD, Jung J, Zhang H, et al. The epidemiology of DSM-5 posttraumatic stress disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. Soc Psychiatry Psychiatr Epidemiol 2016 Aug;51(8):1137-1148. [doi: 10.1007/s00127-016-1208-5] [Medline: 27106853]
13. Gradus JL. Ptsd.va. Epidemiology of PTSD URL: http://www.ptsd.va.gov/professional/PTSD-overview/epidemiological-facts-ptsd.asp [accessed 2017-01-12] [WebCite Cache ID 6nT3D8BQL]
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®), fifth edition. Arlington, VA: American Psychiatric Publishing; 2013.

15. Fischer H. Fas. A guide to U.S. military casualty statistics: operation freedom's sentinel, operation inherent resolve, operation new dawn, operation Iraqi freedom, and operation enduring freedom URL: https://fas.org/sgp/crs/natsec/RS22452.pdf [accessed 2017-10-05] [WebCite Cache ID 6tyvEcpah]
16. Forbes D, Creamer M, Bisson JI, Cohen JA, Crow BE, Foa EB, et al. A guide to guidelines for the treatment of PTSD and related conditions. J Trauma Stress 2010 Oct;23(5):537-552. [doi: 10.1002/jts.20565] [Medline: 20839310]
17. Foa EB, Chrestman KR, Gilboa-Schechtman E. Books.google. 2008. Prolonged exposure therapy for adolescents with PTSD emotional processing of traumatic experiences: therapist guide URL: https://books.google.com/books?hl=en&lr=&id=qlwDAAAQBAJ&oi=fnd&pg=PR11&dq=Foa,+Chrestman,+%26+Gilboa-Schechtman,+2008&ots=3X11fIRO&sig=cUR-TxxKYBIT3qEmVa3SdbOVmg [WebCite Cache ID 6nT3NaVIO]
18. Ptsd.va. Treatment URL: http://www.ptsd.va.gov/PTSD/public/treatment/therapy-med/index.asp [accessed 2017-01-12] [WebCite Cache ID 6nT3WFpOa]
19. Greene-Shortridge TM, Britt TW, Castro CA. The stigma of mental health problems in the military. Mil Med 2007 Feb;172(2):157-161. [Medline: 17357770]
20. Kim PY, Thomas JL, Wilk JE, Castro CA, Hoge CW. Stigma, barriers to care, and use of mental health services among active duty and National Guard soldiers after combat. Psychiatr Serv 2010 Jun;61(6):582-588. [doi: 10.1176/ps.2010.61.6.582] [Medline: 20513681]
21. Mentalhealth.va. VA Mental health services: public report URL: http://www.mentalhealth.va.gov/docs/mental_health_transparency_report_11-24-14.pdf [accessed 2017-01-12] [WebCite Cache ID 6nTF9Johb]
22. Burnam MA, Meredith LS, Tanielian T, Jaycox LH. Mental health care for Iraq and Afghanistan War Veterans. Health Aff (Millwood) 2009;28(3):771-782 [FREE Full text] [doi: 10.1377/hlthaff.28.3.771] [Medline: 19414886]
23. [No authors listed]. Returning home from Iraq and Afghanistan: assessment of readjustment needs of veterans, service members, and their families. Mil Med 2014 Oct;179(10):1053-1055 [FREE Full text] [doi: 10.7205/MILMED-D-14-00263] [Medline: 25269119]
24. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
25. Mittal D, Drummond KL, Blevins D, Curran G, Corrigan P, Sullivan G. Stigma associated with PTSD: perceptions of treatment seeking combat veterans. Psychiatr Rehabil J 2013 Jun;36(2):86-92. [doi: 10.1037/h0094976] [Medline: 23750758]
26. Pietrzak RH, Johnson DC, Goldstein MB, Malley JC, Southwick SM. Perceived stigma and barriers to mental health care utilization among OEF-OIF veterans. Psychiatr Serv 2009 Aug;60(8):1118-1122. [doi: 10.1176/ps.2009.60.8.1118] [Medline: 19648201]
27. American Public Health Association. Apha. 2014. Removing barriers to mental health services for veterans URL: http://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2015/01/28/14/51/removing-barriers-to-mental-health-services-for-veterans [accessed 2017-01-12] [WebCite Cache ID 6nTFiuUTx]
28. Galea S, Basham K, Culpepper L, Davidson J, Foa E, Kizer K, et al. Nationalacademies. 2012. Treatment for posttraumatic stress disorder in military and veteran populations: initial assessment URL: https://nationalacademies.org/hmd/Reports/2012/Treatment-for-Posttraumatic-Stress-Disorder-in-Military-and-Veteran-Populations-Initial-Assessment.aspx [WebCite Cache ID 6nTGN2msW]
29. Smith A. Pewinternet. 2015. U.S. smartphone use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6nTGg6w0B]
30. Sloan DM, Marx BP, Keane TM. Reducing the burden of mental illness in military veterans: commentary on Kazdin and Blase (2011). Perspect Psychol Sci 2011 Sep;6(5):503-506. [doi: 10.1177/1745691611416995] [Medline: 26168203]
31. 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 May 22;2(2):e25 [FREE Full text] [doi: 10.2196/mhealth.2902] [Medline: 25099928]
32. Chan SR, Torous J, Hinton L, Yellowlees P. Mobile tele-mental health: increasing applications and a move to hybrid models of care. Healthcare (Basel) 2014 Jun;2(2):220-233. [Medline: 27429272]
33. Appstore. iTunes Internet store URL: http://appstore.com/ [WebCite Cache ID 6qEtXFlXv]
34. Play.google. Google playstore URL: https://play.google.com/store [WebCite Cache ID 6qEtgXBSa]
35. Mobile.va. VA app store URL: https://mobile.va.gov/appstore [accessed 2017-05-05] [WebCite Cache ID 6qEu7LyDL]
36. t2health.dcoe.mil. Mobile apps URL: http://t2health.dcoe.mil/products/mobile-apps [accessed 2017-05-05] [WebCite Cache ID 6qEu7tP57]
37. Amazon. Appstore for Android URL: http://www.amazon.com/gp/mas/dl/android [WebCite Cache ID 6qEuDqTLd]
38. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
39. Arksey H, O'Malley L. Statistica. 2013. Average selling price for smartphones worldwide in 2013 and 2017, by region (in U.S. dollars) URL: https://www.statista.com/statistics/283334/global-average-selling-price-smartphones/ [WebCite Cache ID 6qKZpRbmi]
40. AT&T. AT&T data plans URL: https://www.att.com/shop/wireless/data-plans.html [accessed 2017-05-09] [WebCite Cache ID 6qKZfJW9c]

41. Verizonwireless. Verizon plan URL: https://www.verizonwireless.com/plans/verizon-plan/ [accessed 2017-05-09] [WebCite Cache ID 6qKZutnvH]
42. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD Coach. JMIR Ment Health 2015 Mar 25;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
43. Kuhn E, Eftekhari A, Hoffman JE, Crowley JJ, Ramsey KM, Reger GM, et al. Clinician perceptions of using a smartphone app with prolonged exposure therapy. Adm Policy Ment Health 2014 Nov;41(6):800-807. [doi: 10.1007/s10488-013-0532-2] [Medline: 24398700]
44. Reger GM, Hoffman J, Riggs D, Rothbaum BO, Ruzek J, Holloway KM, et al. The “PE coach” smartphone application: an innovative approach to improving implementation, fidelity, and homework adherence during prolonged exposure. Psychol Serv 2013 Aug;10(3):342-349. [doi: 10.1037/a0032774] [Medline: 23937084]
45. McInnes DK, Sawh L, Petrakis BA, Rao S, Shimada SL, Eyrich-Garg KM, et al. The potential for health-related uses of mobile phones and internet with homeless veterans: results from a multisite survey. Telemed J E Health 2014 Sep;20(9):801-809. [doi: 10.1089/tmj.2013.0329] [Medline: 25046280]
46. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
47. Gratzer D, Khalid-Khan F, Khalid-Khan S. Anxiety — is there an app for that? Considering technology, psychiatry, and Internet-assisted cognitive behavioral therapy. In: Durbano F, editor. A Fresh Look at Anxiety Disorders. Rijeka, Croatia: InTech; 2015.
48. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
49. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pr 2011 Dec;42(6):505-512. [doi: 10.1037/a0024485]
50. Fletcher RR, Tam S, Omojola O, Redemske R, Kwan J. Wearable sensor platform and mobile application for use in cognitive behavioral therapy for drug addiction and PTSD. Conf Proc IEEE Eng Med Biol Soc 2011;2011:1802-1805. [doi: 10.1109/IEMBS.2011.6090513] [Medline: 22254678]
51. Castro CA. The US framework for understanding, preventing, and caring for the mental health needs of service members who served in combat in Afghanistan and Iraq: a brief review of the issues and the research. Eur J Psychotraumatol 2014 Aug 14;5 [FREE Full text] [doi: 10.3402/ejpt.v5.24713] [Medline: 25206943]
52. Turvey CL, Roberts LJ. Recent developments in the use of online resources and mobile technologies to support mental health care. Int Rev Psychiatry 2015;27(6):547-557. [doi: 10.3109/09540261.2015.1087975] [Medline: 26523397]
53. Kanuri N, Newman MG, Ruzek JI, Kuhn E, Manjula M, Jones M, et al. The feasibility, acceptability, and efficacy of delivering Internet-based self-help and guided self-help interventions for generalized anxiety disorder to Indian university students: design of a randomized controlled trial. JMIR Res Protoc 2015 Dec 11;4(4):e136 [FREE Full text] [doi: 10.2196/resprot.4783] [Medline: 26679295]
54. Baysari MT, Westbrook JI. Mobile applications for patient-centered care coordination: a review of human factors methods applied to their design, development, and evaluation. Yearb Med Inform 2015 Aug 13;10(1):47-54 [FREE Full text] [doi: 10.15265/IY-2015-011] [Medline: 26293851]
55. Mohsenin S, Mohsenin V. Diagnosis and management of sleep disorders in posttraumatic stress disorder:a review of the literature. Prim Care Companion CNS Disord 2014 Dec 11;16(6) [FREE Full text] [doi: 10.4088/PCC.14r01663] [Medline: 25834768]
56. Driesenga SA, Rodriguez JL, Picard T. Evidence-based treatments for military-related posttraumatic stress disorder in a Veterans Affairs setting. Crit Care Nurs Clin North Am 2015 Jun;27(2):247-270. [doi: 10.1016/j.cnc.2015.02.001] [Medline: 25981727]
57. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. Aust N Z J Psychiatry 2013 Feb;47(2):111-113. [doi: 10.1177/0004867412471441] [Medline: 23382507]
58. Price M, Kuhn E, Hoffman JE, Ruzek J, Acierno R. Comparison of the PTSD checklist (PCL) administered via a mobile device relative to a paper form. J Trauma Stress 2015 Oct;28(5):480-483. [doi: 10.1002/jts.22037] [Medline: 26375277]
59. Kuhn E, Crowley JJ, Hoffman J, Eftekhari A, Ramsey KM, Owen JE, et al. Clinician characteristics and perceptions related to use of the PE (prolonged exposure) Coach mobile app. Prof Psychol Res Pr 2015 Dec;46(6):437-443. [doi: 10.1037/pro0000051]
60. Reger GM, Skopp NA, Edwards-Stewart A, Lemus EL. Comparison of prolonged exposure (PE) coach to treatment as usual: a case series with two active duty soldiers. Mil Psychol 2015;27(5):287-296. [doi: 10.1037/mil0000083] [Medline: 26167784]
61. Olff M, Van Zuiden M, Bakker A. Early interventions: from e-health to neurobiology. Eur J Psychotraumatol 2015 Jun 11;6:28545 [FREE Full text] [Medline: 26073209]
62. Weingardt KR, Greene CJ. New electronic tools for veterans. N C Med J 2015;76(5):332-334. [doi: 10.18043/ncm.76.5.332]

63. Christmann CA, Hoffmann A, Bleser G. Stress management apps with regard to emotion-focused coping and behavior change techniques: a content analysis. JMIR Mhealth Uhealth 2017 Feb 23;5(2):e22 [FREE Full text] [doi: 10.2196/mhealth.6471] [Medline: 28232299]
64. Geuens J, Swinnen TW, Westhovens R, de Vlam K, Geurts L, Vanden Abeele V. A review of persuasive principles in mobile apps for chronic arthritis patients: opportunities for improvement. JMIR Mhealth Uhealth 2016 Oct 13;4(4):e118 [FREE Full text] [doi: 10.2196/mhealth.6286] [Medline: 27742604]
65. Taylor DG, Voelker TA, Pentina I. Digitalcommons.sacredheart. 2011. Mobile application adoption by young adults: a social network perspective URL: http://digitalcommons.sacredheart.edu/cgi/viewcontent.cgi?article=1000&context=wcob_fac
66. Cortimiglia MN, Ghezzi A, Renga F. Mobile applications and their delivery platforms. IT Prof 2011;13(5):51-56. [doi: 10.1109/MITP.2011.84]
67. Lymberis A. Smart wearables for remote health monitoring, from prevention to rehabilitation: current R&D, future challenges. 2003 Presented at: 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine, 2003; 2003; Birmingham p. 272-275. [doi: 10.1109/ITAB.2003.1222530]
68. Spoont MR, Murdoch M, Hodges J, Nugent S. Treatment receipt by veterans after a PTSD diagnosis in PTSD, mental health, or general medical clinics. Psychiatr Serv 2010 Jan;61(1):58-63. [doi: 10.1176/appi.ps.61.1.58] [Medline: 20044419]
69. Alghowinem S, Goecke R, Wagner M, Epps J, Breakspear M, Parker G. Citeseerx.ist.psu. 2012. From joyous to clinically depressed: mood detection using spontaneous speech URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.655.4340&rep=rep1&type=pdf [WebCite Cache ID 6nTKTD09T]
70. Berry ME, Chapple IT, Ginsberg JP, Gleichauf KJ, Meyer JA, Nagpal ML. Non-pharmacological intervention for chronic pain in veterans: a pilot study of heart rate variability biofeedback. Glob Adv Health Med 2014 Mar;3(2):28-33 [FREE Full text] [doi: 10.7453/gahmj.2013.075] [Medline: 24808979]

Abbreviations

ACT: Acceptance and Commitment Therapy
CBT: cognitive behavioral therapy
CPT: cognitive processing therapy
EBT: evidence-based treatment
HFE: human factors and ergonomics
mHealth: mobile health
PE: prolonged exposure therapy
PTSD: posttraumatic stress disorder
VA: Veterans Affairs

Edited by G Eysenbach; submitted 12.01.17; peer-reviewed by S Chan, K Weingardt; comments to author 22.03.17; revised version received 05.08.17; accepted 29.08.17; published 26.10.17.

Please cite as:
Rodriguez-Paras C, Tippey K, Brown E, Sasangohar F, Creech S, Kum HC, Lawley M, Benzer JK
Posttraumatic Stress Disorder and Mobile Health: App Investigation and Scoping Literature Review
JMIR Mhealth Uhealth 2017;5(10):e156
URL: http://mhealth.jmir.org/2017/10/e156/
doi:10.2196/mhealth.7318
PMID:29074470

Original Paper

Impact of a Telehealth Program With Voice Recognition Technology in Patients With Chronic Heart Failure: Feasibility Study

Heesun Lee[1,2], MD; Jun-Bean Park[1,3], MD, PhD; Sae Won Choi[4], MD; Yeonyee E Yoon[1,5], MD; Hyo Eun Park[1,2], MD; Sang Eun Lee[6], MD, PhD; Seung-Pyo Lee[1,3], MD, PhD; Hyung-Kwan Kim[1,3], MD, PhD; Hyun-Jai Cho[1,3], MD, PhD; Su-Yeon Choi[1,2], MD, PhD; Hae-Young Lee[1,3], MD, PhD; Jonghyuk Choi[7], MD; Young-Joon Lee[7], MD; Yong-Jin Kim[1,3], MD, PhD; Goo-Yeong Cho[1,5], MD, PhD; Jinwook Choi[8], MD, PhD; Dae-Won Sohn[1,3], MD, PhD

[1]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic Of Korea

[2]Healthcare System Gangnam Center, Seoul National University Hospital, Seoul, Republic Of Korea

[3]Cardiovascular Center, Seoul National University Hospital, Seoul, Republic Of Korea

[4]Office of Hospital Information, Seoul National University Hospital, Seoul, Republic Of Korea

[5]Cardiovascular Center, Seoul National University Bundang Hospital, Seongnam, Republic Of Korea

[6]Department of Cardiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic Of Korea

[7]AIMMED Co., Ltd., Seoul, Republic Of Korea

[8]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Republic Of Korea

Corresponding Author:
Jun-Bean Park, MD, PhD
Cardiovascular Center
Seoul National University Hospital
101 Daehak-ro, Jongro-gu
Seoul, 110-744
Republic Of Korea
Phone: 82 2 2072 2252
Fax: 82 2 2072 4922
Email: nanumy1@gmail.com

Abstract

Background: Despite the advances in the diagnosis and treatment of heart failure (HF), the current hospital-oriented framework for HF management does not appear to be sufficient to maintain the stability of HF patients in the long term. The importance of self-care management is increasingly being emphasized as a promising long-term treatment strategy for patients with chronic HF.

Objective: The objective of this study was to evaluate whether a new information communication technology (ICT)–based telehealth program with voice recognition technology could improve clinical or laboratory outcomes in HF patients.

Methods: In this prospective single-arm pilot study, we recruited 31 consecutive patients with chronic HF who were referred to our institute. An ICT-based telehealth program with voice recognition technology was developed and used by patients with HF for 12 weeks. Patients were educated on the use of this program via mobile phone, landline, or the Internet for the purpose of improving communication and data collection. Using these systems, we collected comprehensive data elements related to the risk of HF self-care management such as weight, diet, exercise, medication adherence, overall symptom change, and home blood pressure. The study endpoints were the changes observed in urine sodium concentration (uNa), Minnesota Living with Heart Failure (MLHFQ) scores, 6-min walk test, and N-terminal prohormone of brain natriuretic peptide (NT-proBNP) as surrogate markers for appropriate HF management.

Results: Among the 31 enrolled patients, 27 (87%) patients completed the study, and 10 (10/27, 37%) showed good adherence to ICT-based telehealth program with voice recognition technology, which was defined as the use of the program for 100 times or more during the study period. Nearly three-fourths of the patients had been hospitalized at least once because of HF before the enrollment (20/27, 74%); 14 patients had 1, 2 patients had 2, and 4 patients had 3 or more previous HF hospitalizations. In the total study population, there was no significant interval change in laboratory and functional outcome variables after 12 weeks

of ICT-based telehealth program. In patients with good adherence to ICT-based telehealth program, there was a significant improvement in the mean uNa (103.1 to 78.1; *P*=.01) but not in those without (85.4 to 96.9; *P*=.49). Similarly, a marginal improvement in MLHFQ scores was only observed in patients with good adherence (27.5 to 21.4; *P*=.08) but not in their counterparts (19.0 to 19.7; *P*=.73). The mean 6-min walk distance and NT-proBNP were not significantly increased in patients regardless of their adherence.

Conclusions: Short-term application of ICT-based telehealth program with voice recognition technology showed the potential to improve uNa values and MLHFQ scores in HF patients, suggesting that better control of sodium intake and greater quality of life can be achieved by this program.

(JMIR Mhealth Uhealth 2017;5(10):e127) doi:10.2196/mhealth.7058

KEYWORDS

heart failure; telemedicine; selfcare; compliance

Introduction

Heart failure (HF) is a major public health issue, with a prevalence of over 23 million worldwide and rising with the aging of the population [1,2]. The characteristic feature of HF is a progressive loss of cardiomyocytes and the development of cardiac dysfunction, ultimately leading to frequent hospitalization and significant morbidity and mortality [1,3-5]. There are ongoing efforts to develop improved therapeutic and preventive strategies for HF, as most current strategies fail to reduce readmission rates and maintain the stability of HF patients in the long term [6,7].

Most hospitalizations for acute decompensated HF are attributable to poor self-care, including lack of knowledge, nonadherence to proper diet or medications, and failure of self-management of symptoms. Recent studies on patient-centered out-of-hospital management of HF patients have shown improvements in prognosis and quality of life [3,8-10]. However, in practice, barriers to self-care management such as insufficient personalized real-time feedback to patients regarding body weight, fluid balance, and physical activity limit their effectiveness [8,11,12]. Frequent assessment of physiologic parameters related to HF aggravation and remote disease management aided by advances in telemonitoring systems based on information communication technology (ICT) may be an approach to overcome the limitations of self-management. By enabling early detection of HF decompensation and through timely intervention, it may be possible to achieve improved outcomes and reduced medical costs for HF patients [11,13].

Voice recognition technology enables the recognition and translation of verbal information into text, which can then be used in automated data processing systems. With advances in the accuracy of this technology, it is being applied in the medical field to facilitate interest and improve the adherence of patients with chronic conditions such as asthma, diabetes mellitus, and glaucoma [14-16]. In a recent study, an ICT-based telehealth program with voice recognition technology was effective in achieving glycemic control without hypoglycemia in elderly diabetic patients [15]. Given that voice recognition technology can facilitate the gathering of data that are currently difficult to retrieve, as it enables patients, particularly the elderly, to easily access the system and input their data and does not entail additional equipment costs to collect the necessary data, an ICT-based telehealth program incorporating this technology may be a promising tool for improving clinical or laboratory outcomes in chronic HF patients. However, data regarding the role of this novel suite of technologies in the self-management of HF patients are scarce.

With the expectation of enhanced interactivity and active engagement of participants and delivery of tailored interventions, ultimately leading to effective self-care of HF, we developed an ICT-based telehealth program featuring voice recognition technology for HF and sought to explore whether this new program influences the control of sodium intake and quality of life in patients with chronic HF.

Methods

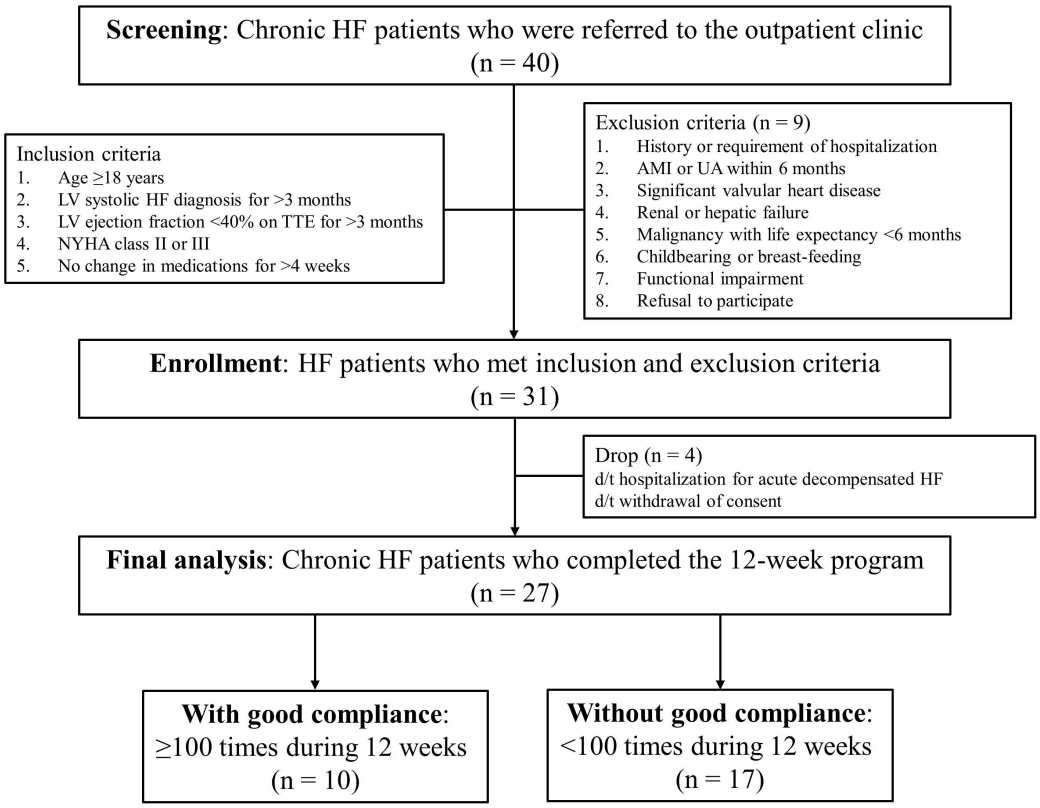
Study Design and Population

This study was designed as a prospective single-arm pilot study. The study sample was recruited from consecutive HF patients who were referred to the outpatient clinic of our institute since July 2014. Inclusion criteria were as follows: (1) age ≥18 years; (2) left ventricular (LV) systolic HF diagnosis for >3 months, regardless of etiology according to current guidelines [3,9]; (3) LV ejection fraction (EF) <40% on transthoracic echocardiography of a duration of at least 3 months; (4) New York Heart Association (NYHA) class II or III; and (5) clinical stability including no change in medications for at least the past 4 weeks. Exclusion criteria were the following: (1) hospitalization within 3 months or a condition expected to require hospitalization; (2) a history of acute myocardial infarction or unstable angina within 6 months; (3) a history of significant valvular heart disease requiring surgical or interventional correction; (4) renal replacement therapy; (5) severe hepatic dysfunction defined as aspartate transaminase and alanine transaminase ≥3 times the upper normal limits; (6) malignancy with life expectancy <6 months; (7) childbearing potential or breastfeeding; (8) functional impairment related to severe musculoskeletal or neurological problems; and (9) refusal to participate. The study protocol was in accordance with the Declaration of Helsinki and approved by the institutional review board of our institution. Written informed consent was obtained from all the enrolled subjects.

Patients were educated regarding the full range of HF self-care behaviors, including diet, exercise, and medication adherence. All participants underwent the following assessments at baseline and after 12 weeks of the ICT-based telehealth program: medical

history, physical examination, laboratory tests including urine sodium concentration (uNa) and N-terminal prohormone of brain natriuretic peptide (NT-proBNP), 6-min walk test, NYHA functional class, and Minnesota Living with Heart Failure Questionnaire (MLHFQ) scores [17-19]. The study endpoints included the changes in uNa, MLHFQ scores, NT-proBNP, and 6-min walk distance after 12 weeks of intervention. The adherence to the telehealth program was also analyzed, and a good adherence was defined as the use of the program for 100 times or more during the 12-week intervention period. The study flow is illustrated in Figure 1.

Figure 1. Study flow of ICT-based telehealth program in HF. ICT: information communication technology. HF: heart failure, LV: left ventricular, NYHA: New York Heart Association, TTE: transthoracic echocardiography, AMI: acute myocardial infarction, UA: unstable angina.



The ICT-Based Telehealth Program

The ICT-based telehealth center comprised multidisciplinary team members, including experienced cardiologists, professional nurses, a dietitian, and computer programmers. Participants were instructed to use the ICT-based telehealth program with voice recognition technology (AIMMED Co Ltd) that was developed for HF patients to provide data and to communicate with clinicians more conveniently through ordinary communication channels via their handheld mobile phone or landline. Patients used their voice or touch-tone telephone keypads to enter essential information for the assessment of HF, including daily weight, diet, exercise, medication adherence, overall symptom change, and home blood pressure. Figure 2 illustrates the schematic diagram of the ICT-based telehealth program.

Direct voice input of health-related data was performed via landline or mobile phone through the automatic call and response system. The automatic call system was programmed to automatically dial the participants' registered phone numbers once daily 5 days per week for 12 weeks, and when a person answered the phone, it asked prespecified questions to collect narrative responses. Additionally, patients were also allowed free calls into the system during the study period to collect health-related data from participants who did not answer the call from the automatic call system. Patients, guided by a series of prespecified verbal questions, were requested to speak or to enter a preset range of inputs, including body weight, blood pressure, heart rate, and HF-related symptoms. Examples of possible answers were also presented for patients' convenience and to reduce the error of the data collection. To further ensure the accuracy of data, the program provided patients the opportunity to either confirm or correct the initial extracted data. An example flow diagram for voice recognition is shown in Figure 3.

A clinical decision support system was implemented to provide immediate tailored feedback by voice or text messages (short message service, SMS) to patients according to a predetermined algorithm following current guidelines [3,20]. When data entries were outside predefined ranges or symptoms were reported, an automatic phone call was made to patients to confirm the data and to their attending physicians and research nurses to notify them of the occurrence of these events. Patients received timely self-care feedback from their physicians and were monitored to ensure resolution. Attending physicians were able to tailor HF management and adjust clinic visit schedules, taking into account the clinical status of each patient. After an out-of-range value occurred, the choice of whether to modify the limits of

the range was at the discretion of the physician, who made the decision weighing the individual risk and benefit. Patients were instructed on the use of the website or mobile app created for this study, which provided individually tailored feedback messages based on data entered by the patients (Multimedia Appendix 1). By analyzing the cumulative data of each individual, patients also received personalized guidance on long-term self-care management of HF via the website or as a mobile report. We summarized the functionalities of the ICT-based telehealth program used in our study compared with previous programs in Multimedia Appendix 2.

Figure 2. Schematic diagram of the ICT-based telehealth program in HF. ICT: information communication technology, HF: heart failure, TTS: text to speech, IVR: interactive voice response, PBX: private branch exchange, CTI: computer telephony integration, TCP: transmission control protocol, WAS: Web application server, REC: recording, IP: Internet Protocol.

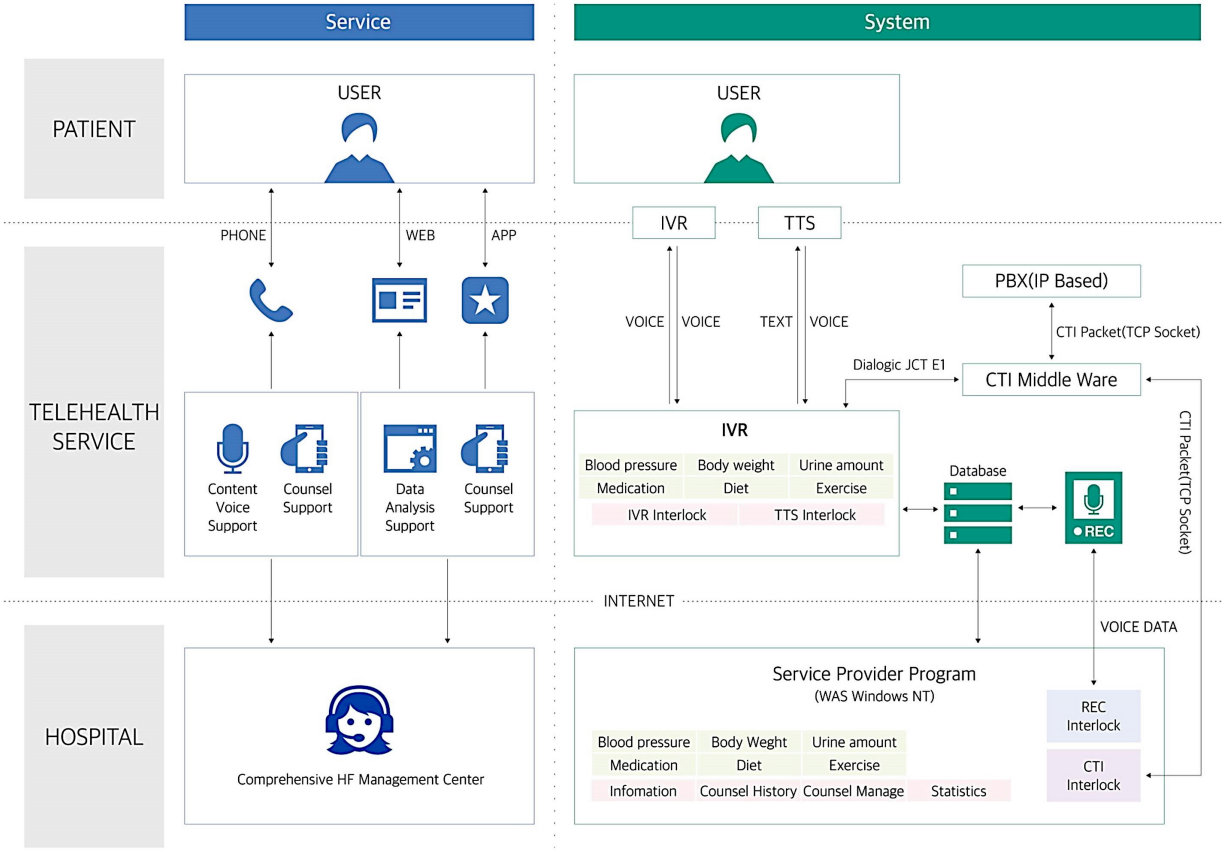
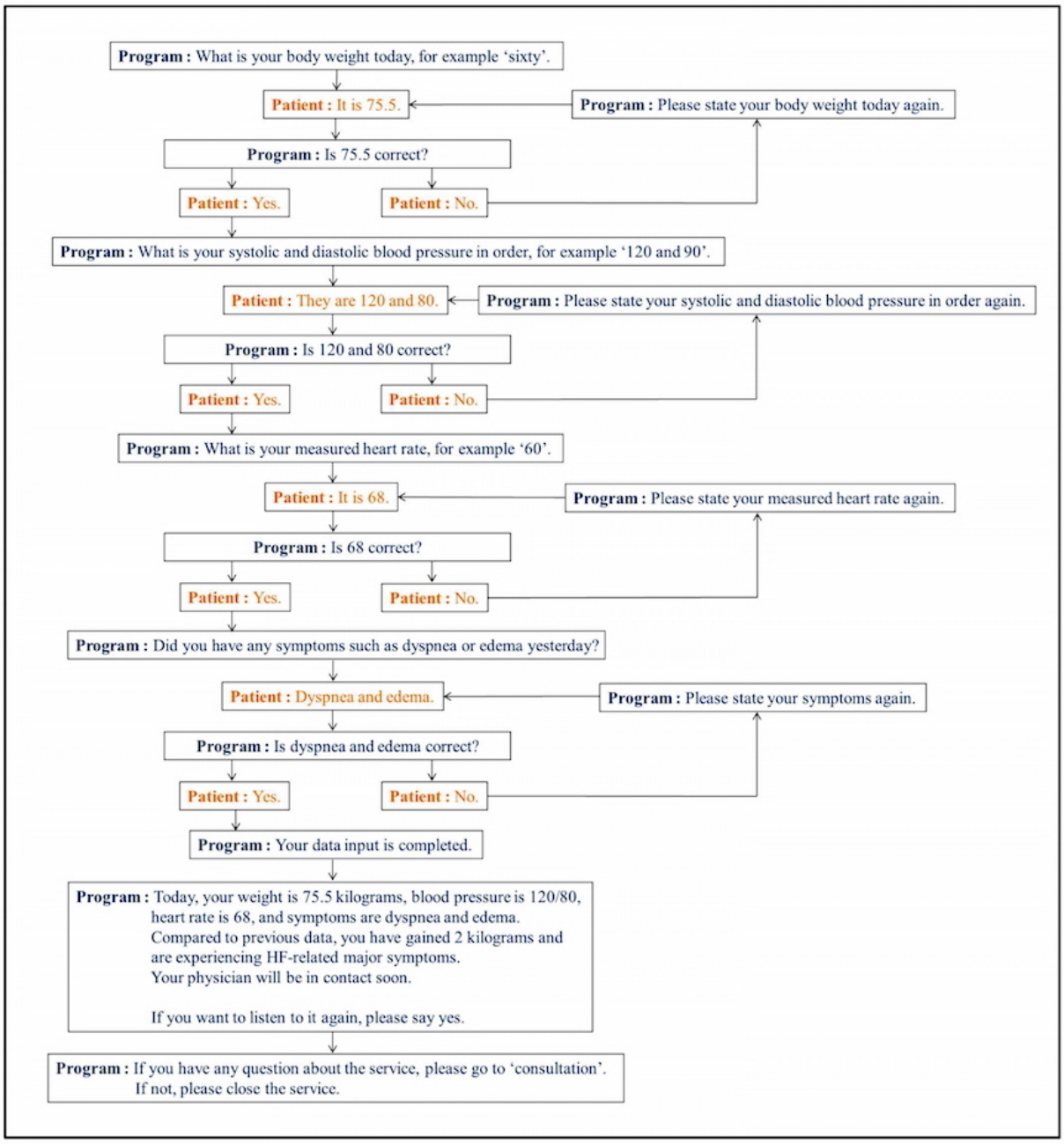


Figure 3. An example of flow diagram for voice recognition program. The voice recognition component of the automatic call and response system enabled the recognition of the patient's voice data to be input into the ICT-based telehealth program. Patients were requested to speak a pre-set range of inputs, such as body weight, blood pressure, heart rate, and heart failure-related symptoms. Patients were guided by a series of pre-specified verbal questions and examples of possible answers. To ensure the accuracy of the obtained data, the program allowed patients to either confirm or correct the initial extracted data.



Data Collection

Venous blood and urinary samples were obtained from all patients at baseline and after completion of 12 weeks of ICT-based telehealth program. Data including uNa, NT-proBNP, 6-min walk test, and MLHFQ scores were collected to assess the impact of this program on patient outcomes. Specifically, uNa and NT-proBNP were used to assess the adherence to a low-salt diet and the progression of HF, respectively. To assess functional capacity, 6-min walk test was conducted using a standardized protocol between 11 AM to 2 PM on usual medications [21,22]. Briefly, patients were required to perform a 6-min shuttle walk test with markers placed every 25 m. During the 6-min walk test, patients were accompanied by trained research nurses, who were independent of this study, for the safety of participants as well as for the accuracy and reliability of the results. We also used MLHFQ scores to quantify and compare health-related quality of life outcomes before and after the telehealth program [17-19]. The MLHFQ is an established questionnaire, which was developed to measure the subjective influence of HF and HF treatments on an individual's quality of life, including physical, socioeconomic, and emotional aspects. It comprises 21 items with scores ranging from 0 to 105, and higher scores indicate a lower quality of life in HF patients [17,19].

Statistical Analysis

Descriptive statistics were used to characterize the study population and measurements. Data were presented as numbers and percentages for categorical variables and mean ± standard error for continuous variables. Differences between continuous variables were compared by the Student *t* test or Mann-Whitney *U* test for independent samples, and those between categorical variables were analyzed by the chi-square test or Fisher exact test, as appropriate. A *P* value of <.05 was considered statistically significant. The Wilcoxon signed-rank test was used to compare values before and after ICT-based telehealth program. In addition, subgroup analysis was also performed according to the adherence to intervention.

Results

Baseline Characteristics of Study Population

Of the 40 chronic HF patients who were screened, 31 subjects met the inclusion and exclusion criteria and agreed to participate in this study. Furthermore, 4 patients (4/31, 13%) were dropped from the study because of hospitalization for acute decompensated HF (n=2) and withdrawal of consent (n=2). Two patients hospitalized for acute decompensated HF withdrew within 1 week and 4 weeks of starting the study, respectively, and the cause of HF hospitalization was considered to be a viral infection in both patients. The final analysis was performed in 27 patients (27/31, 87%) who completed the baseline and 12-week measurements. Ten patients (10/27, 37%) showed good adherence to the program, defined as the frequency of its use ≥100 times during the 12-week intervention period. Table 1 describes the baseline characteristics of the total study population and subgroups according to adherence to the intervention. The mean age of the total study population was 63.4 years, and one-third were females (9/27, 33%). Nearly three-fourths had been hospitalized at least once because of HF before the enrollment (20/27, 74%). The number of prior HF hospitalizations per individual was 1.15; 14 patients had 1, 2 patients had 2, and 4 patients had 3 or more previous HF hospitalizations. Before the program started, 70% of patients were already undergoing treatment with diuretics, including mineralocorticoid receptor antagonists (19/27, 70%). There were no changes in HF medications or dosages during the study period. The etiology of 8 patients was ischemia. The baseline mean values of LV ejection fraction, MLHFQ scores, and 6-min walk distance were 30%, 22.2 points, and 408.4 m, respectively. When comparing the baseline characteristics between patients according to adherence to the program, there was no significant difference between the groups. Patients with good adherence tended to have worse baseline parameters, including significantly more frequent hospitalizations related to HF, greater levels of NT-proBNP and uNa, higher MLHFQ scores, and shorter 6-min walk distances, compared with those without (Table 1).

Approach and Satisfaction of ICT-Based Telehealth Program

Twenty-one patients utilized the program via their mobile phone or landline, and about 59% (16/27) of them also accessed the website. The accuracy of voice recognition was 93% and 95% in patients using mobile phone and landline, respectively. Of the patients in whom the user experience with the voice recognition system could be evaluated, ≥85% patients (23/27) indicated that they were satisfied or neutral with the service provided, as shown in Multimedia Appendix 3. Patients were interested in the following services: blood pressure (27/27, 100%), followed by body weight (26/27, 96%; 951 cases/week), and medication (26/27, 96%; 776 cases/week). Table 2 shows the utilization according to the method and the components accessed.

Table 1. Baseline characteristics according to adherence to information communication technology (ICT)–based telehealth program in heart failure (HF).

Characteristics	All patients (N=27)	Patients with good adherence (n=10)	Patients without good adherence (n=17)	*P* values
Demographics				
Age in years, mean ± SE[a]	63.4 ± 1.8	62.1 ± 2.7	64.1 ± 2.4	.60
Female gender, n (%)	9 (33)	2 (20)	7 (41)	.41
Height in cm, mean ± SE	161.4 ± 1.4	164.7 ± 1.9	159.4 ± 1.8	.07
Weight in kg, mean ± SE	63.7 ± 1.9	64.8 ± 2.8	63.1 ± 2.7	.70
Body mass index in kg/m^2, mean ± SE	24.4 ± 0.6	23.9 ± 0.4	24.7 ± 0.6	.52
Clinical history, n (%)				
Hypertension	12 (44)	4 (40)	8 (47)	.72
Diabetes mellitus	6 (22)	3 (30)	3 (18)	.46
Atrial fibrillation	11 (41)	4 (40)	7 (41)	.95
Chronic obstructive pulmonary disease	1 (4)	1 (10)	0 (0)	.18
Chronic kidney disease	10 (37)	4 (40)	6 (35)	.81
Prior myocardial infarction	9 (33)	5 (51)	4 (24)	.16
Prior hospitalization due to HF[b]	20 (74)	6 (60)	14 (82)	.20
Signs and symptoms on enrollment, mean ± SE				
Systolic BP[c], mm Hg	117.3 ± 2.5	116.1 ± 4.3	117.9 ± 3.1	.73
Diastolic BP, mm Hg	68.4 ± 1.3	68.5 ± 1.4	68.4 ± 1.9	.98
Heart rate, beats per minute	67.9 ± 2.3	64.4 ± 4.0	69.9 ± 2.8	.26
Medications on enrollment, n (%)				
Diuretics	19 (70)	6 (60)	13 (77)	.42
ACEI[d]	9 (33)	4 (40)	4 (29)	.68
ARB[e]	15 (56)	4 (40)	11 (65)	.26
Beta blocker	12 (44)	4 (40)	8 (47)	>.99
Calcium channel blocker	4 (15)	1 (10)	3 (18)	>.99
Spironolactone	19 (70)	7 (70)	12 (71)	>.99
Digoxin	8 (30)	3 (30)	5 (29)	>.99
Etiology, n (%)				
Ischemic	8 (30)	4 (40)	4 (29)	.37
Nonischemic	19 (70)	6 (60)	13 (77)	.42
Laboratory variables, mean ± SE				
WBC[f], × 10^3/uL	6.4 ± 0.3	6.1 ± 0.5	6.6 ± 0.3	.37
Hemoglobin, g/dL	13.4 ± 0.3	13.6 ± 0.5	13.3 ± 0.4	.60
Platelet, × 10^3/uL	196.7 ± 9.5	178.6 ± 14.3	207.3 ± 12.0	.15
Sodium, mmol/L	139.6 ± 0.5	140.2 ± 0.7	139.2 ± 0.6	.35
Potassium, mmol/L	4.6 ± 0.1	4.7 ± 0.2	4.5 ± 0.1	.27
Chloride, mmol/L	102.9 ± 0.6	104.6 ± 0.8	101.9 ± 0.8	.13
Total CO_2[g], mmol/L	28.5 ± 0.6	27.2 ± 1.2	29.2 ± 0.7	.13

Characteristics		All patients (N=27)	Patients with good adherence (n=10)	Patients without good adherence (n=17)	*P* values
	Calcium, mg/dL	9.4 ± 0.1	9.2 ± 0.2	9.5 ± 0.1	.13
	Phosphate, mg/dL	3.6 ± 0.1	3.5 ± 0.1	3.8 ± 0.1	.13
	Blood urea nitrogen, mg/dL	18.2 ± 1.3	17.7 ± 2.5	18.5 ± 1.5	.78
	Creatinine, mg/dL	1.1 ± 0.1	1.2 ± 0.3	1.0 ± 0.1	.35
	GFR[h], mL/min/1.73 m^2	69.8 ± 3.6	70.6 ± 7.7	69.3 ± 3.8	.88
	NT-proBNP[i], pg/mL	288.6 ± 65.7	300.2 ± 108.1	281.7 ± 85.2	.90
	Urine sodium, mmol/L	91.9 ± 0.5	103.1 ± 9.2	85.4 ± 10.5	.26
	LV[j] ejection fraction, %	30.3 ± 1.3	30.7 ± 2.4	30.1 ± 1.5	.81
Quality of life data, mean ± SE	MLHFQ[k] score	22.2 ± 3.2	27.5 ± 6.4	19.0 ± 3.3	.20
Functional status data, mean ± SE	6-min walk distance, m	408.4 ± 16.1	404.0 ± 30.1	411.0 ± 19.2	.84

[a]SE: standard error.

[b]HF: heart failure.

[c]BP: blood pressure.

[d]ACEI: angiotensin-converting enzyme inhibitor.

[e]ARB: angiotensin receptor blocker.

[f]WBC: white blood cell.

[g]CO_2: carbon dioxide.

[h]GFR: glomerular filtration rate.

[i]NT-proBNP: N-terminal prohormone of brain natriuretic peptide.

[j]LV: left ventricular.

[k]MLHFQ: Minnesota Living with Heart Failure Questionnaire.

Table 2. Data on utilization of information communication technology (ICT)–based telehealth program in heart failure (HF).

Data on utilization		Patients accessed, n (%)	Frequency of access per week
Type of access method			
	Mobile phone or landline	21 (77.8)	2297
	ACS[a]/ARS[b]	10 (37.0)	139
	TTS[c]	18 (66.7)	848
	Website access	16 (59.3)	433
Type of accessed component			
	Body weight	26 (96.3)	951
	Diet	24 (88.9)	419
	Exercise	25 (92.6)	657
	Medication	26 (96.3)	776
	Overall symptom change	22 (81.5)	651
	Blood pressure	27 (100.0)	2812

[a]ACS: automatic call system.

[b]ARS: automatic response system.

[c]TTS: text to speech.

Changes After Using ICT-Based Telehealth Program

In the total study population, there was no significant interval change in laboratory and functional outcome variables after 12 weeks of ICT-based telehealth program (Figure 4): uNa (91.9 to 90.0; *P*=.45), MLHFQ scores (22.2 to 20.3; *P*=.18), 6-min walk distance, (408.4 to 410.6; *P*=.59), and NT-proBNP (288.6 to 318.1; *P*=.91). When we stratified patients into two groups by adherence to the ICT-based telehealth program, there was a significant reduction in the value of uNa (103.1 to 78.1; *P*=.01) and a marginal decrease in the MLHFQ scores (27.5 to 21.4; *P*=.08) in HF patients with good adherence (Figure 5). However, no significant changes were observed in those without good adherence regarding these variables (Figure 6). On the other hand, 6-min walk distance neither significantly improved in patients with good adherence nor in those without (Figure 6). No significant change in NT-proBNP was observed in both groups of patients (Figures 5 and 6). The MLHFQ scores decreased after 12 weeks of ICT-based telehealth program in 18 patients (18/27, 67%). When we assessed the change in the subscores of physical (8 items), emotional (5 items), and socioeconomic (8 items) dimensions of the MLHFQ, although statistically insignificant, patients with good adherence showed numerically greater improvements in all dimensions. Specifically, the improvement in the total score of MLHFQ was largely driven by the decrease in emotional dimension measures among three quality of life dimensions (94%, 17/18 patients). The change in emotional domain scores was −2.1 in patients with good adherence and −0.2 in those without (*P*=.48). Results in terms of physical dimension measures revealed that the change was −2.5 in patients with good adherence and 0.35 in those without (*P*=.13). Regarding socioeconomic dimension measures, the change was −1.5 in patients with good adherence and 0.5 in those without (*P*=.11).

Figure 4. Changes in major outcomes over time by ICT-based telehealth program in total study population. There was no significant changes in uNa (a), MLHFQ scores (b), 6-minute walk distance (c), and NT-proBNP level (d) after 12 weeks of ICT-based telehealth program. ICT: information communication technology, HF: heart failure, NT-proBNP: N-terminal prohormone of brain natriuretic peptide, MLHFQ: Minnesota Living with Heart Failure Questionnaire.

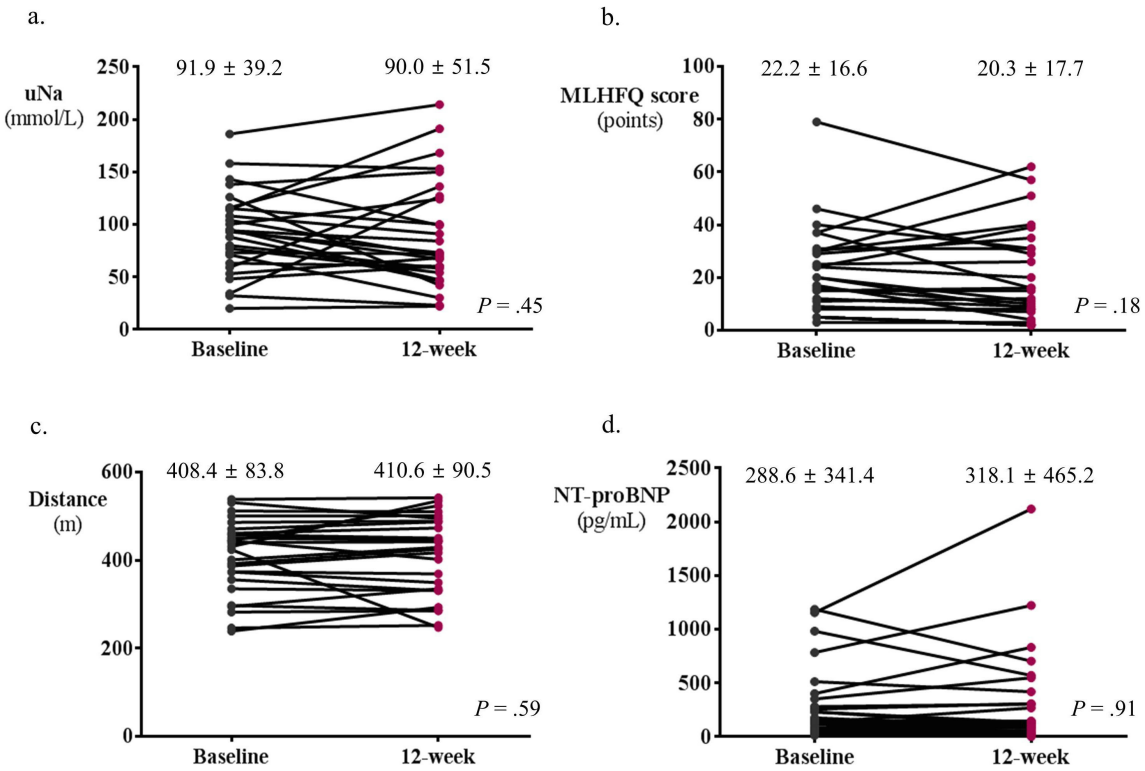


Figure 5. Changes in major outcomes over time by information communication technology (ICT)–based telehealth program in patients with good adherence. In patients with good adherence, uNa significantly decreased after 12 weeks of intervention (a), whereas MLHFQ scores marginally decreased (b). There were no changes in 6-minute walk distance (c) and NT-proBNP level (d). NT-proBNP: N-terminal prohormone of brain natriuretic peptide, MLHFQ: Minnesota Living with Heart Failure Questionnaire.

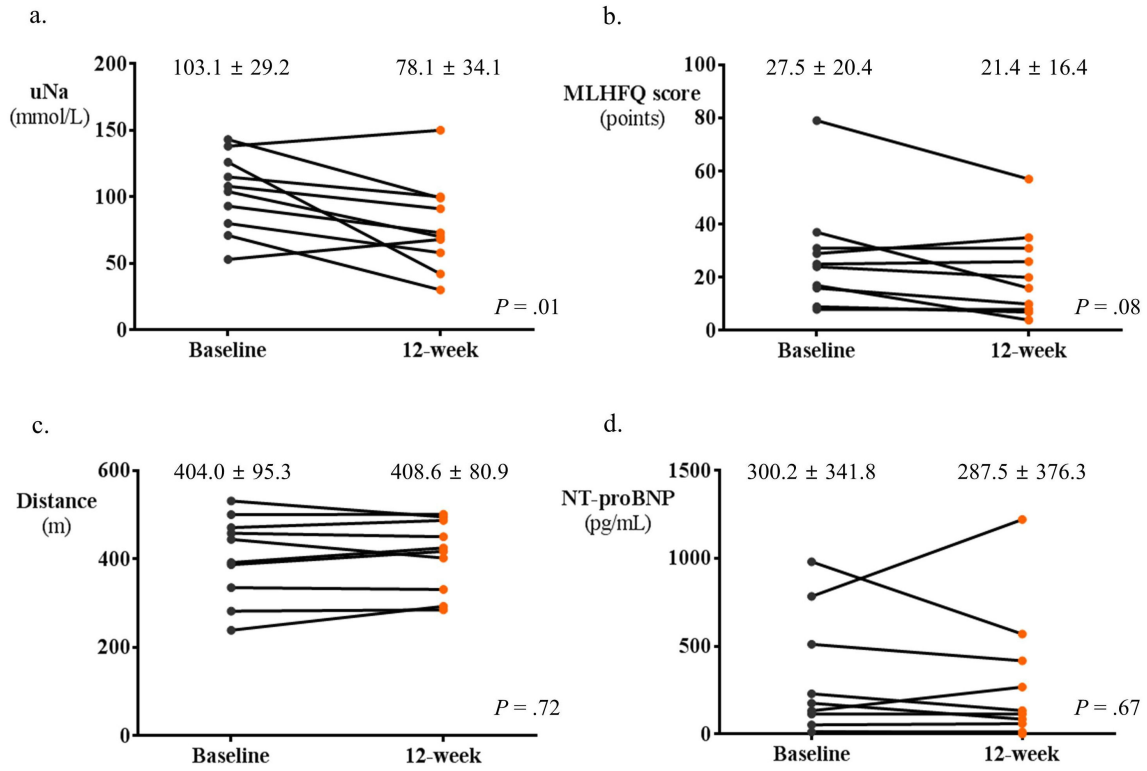
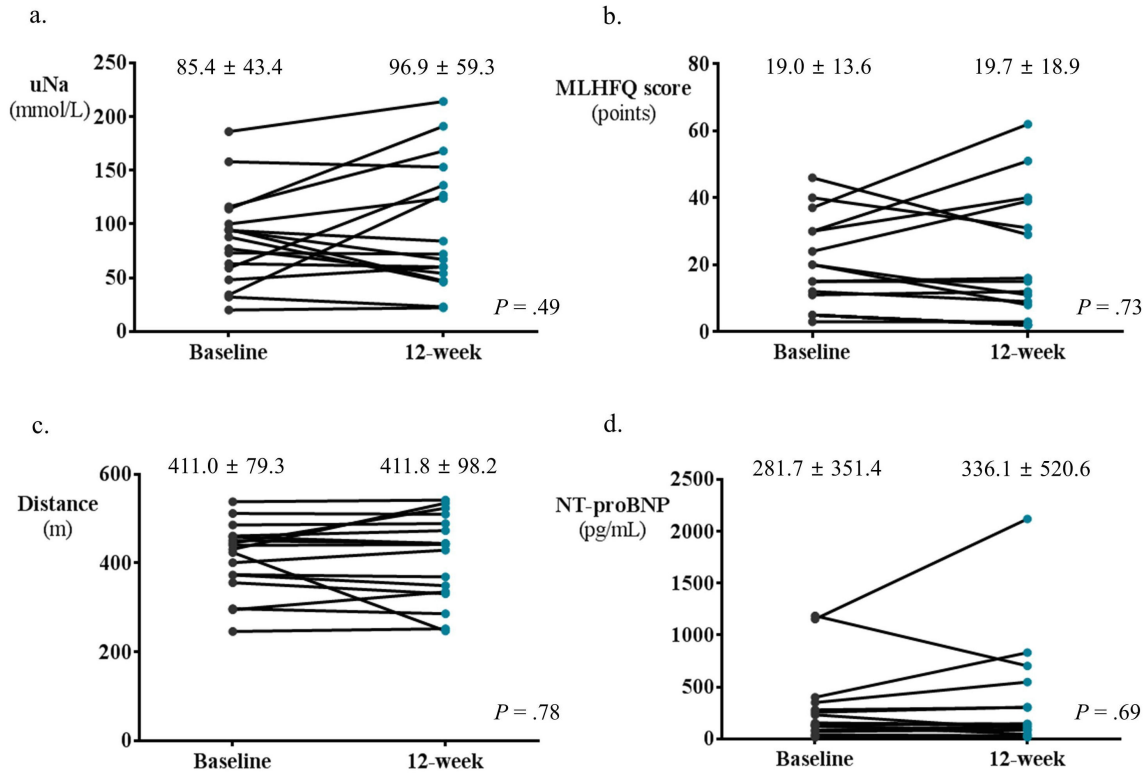


Figure 6. Changes in major outcomes over time by information communication technology (ICT)–based telehealth program in patients without good adherence. In patients without good adherence, no significant changes were observed in laboratory and functional outcomes after 12 weeks of ICT-based telehealth program (a-d). NT-proBNP: N-terminal prohormone of brain natriuretic peptide, MLHFQ: Minnesota Living with Heart Failure Questionnaire.



Discussion

Principal Findings

This study is a prospective pilot study examining the possibility of ICT-based telehealth program with voice recognition technology in HF and building up the foundation for further large-scale randomized controlled studies. Our study demonstrated that chronic HF patients with good adherence to ICT-based telehealth program showed improvement in uNa and in symptoms evaluated by the MLHFQ scores. To the best of our knowledge, this is the first study to implement this novel technology to improve self-care of HF. Our study may pave the way for further large-scale clinical outcome trials to determine the efficacy and feasibility of ICT-based interventions for long-term management of HF patients.

Impact of ICT-Based Telehealth Program in HF on Urine Sodium Concentration

The key characteristics of HF include an increase in sodium avidity and a tendency toward congestion [23,24]. Therefore, the maintenance of sodium and water balance is a critical element of HF management, and among several noninvasive methods proposed to assess this balance, uNa has been increasingly used as a simple, inexpensive, and noninvasive indicator for the dietary sodium intake [25-29]. Specifically, it is well established that the urinary sodium excretion is precisely regulated to match dietary sodium intake [25,30]. Although the assessment of 24-hour urine sodium excretion is considered the gold standard method, the measurement of spot uNa could provide an approximate estimate of sodium excretion, with increased convenience and low cost [25,29,31-34]. On the other hand, a previous study demonstrated that spot uNa was reduced by decongestive therapy for advanced HF, which might provide insightful information to titrate the diuretic dose [35,36]. Hence, in general, decreased uNa in HF patients can be a result of low salt intake or effective diuretic therapy [35]. In this study, HF patients with good adherence to ICT-based telehealth program showed reduced uNa after 12 weeks of intervention. Given that patients were clinically stable without any evidence of acute destabilized HF and change in diuretic dose, decreased uNa might be a reflection of better control of dietary salt intake. As recent data emphasize nonpharmacological as well as pharmacological interventions for HF because of its progressive nature [37], an ICT-based telehealth program can be a useful tool for improving self-care and, potentially, prognosis in HF patients [37].

Impact of ICT-Based Telehealth Program in HF on the MLHFQ Scores

Improving health-related quality of life is not only a major treatment goal in HF but also an effective surrogate marker of appropriate HF therapy and improved prognosis in HF patients [17,38-40]. Among various methods of assessing health-related quality of life, we used the MLHFQ, which has been extensively validated in HF patients [18,19,38,40,41]. In our study, the mean MLHFQ scores numerically decreased by about 2 points in all patients (P=.18) and by 6.1 points in patients with good adherence (P=.08) after 12 weeks of ICT-based telehealth program. Although there is no absolute cutoff for the MLHFQ scores to discriminate between normal and impaired quality of life, a decrease of more than 5 points is considered as a clinically important change [42,43]. Previous studies have suggested that the changes in the MLHFQ scores can translate into differences in clinical outcomes of HF patients. For example, EPICAL trial showed that a 10-point decrement in the MLHFQ scores was associated with a 23% decrease in mortality and a 31% decrease in readmission for HF [17]. Hoekstra et al [42] also reported that each 10-point increment in the MLHFQ scores was associated with a 7% increase in mortality. Given that the short duration of intervention in this study led to a 6.1 point decrease in the MLHFQ scores among patients with good adherence, an ICT-based telehealth program seems promising to enhance quality of life and consequently survival of HF patients. Since the MLHFQ comprises three subscores (physical, emotional, and socioeconomic dimensions), we also compared the changes in each of the subscores between groups. In this analysis, the reduction in the MLHFQ scores mostly stemmed from the improvement in emotional aspects in patients with good adherence, although the between-group difference was not significant. These findings suggest that intense monitoring and automated feedback using an ICT-based telehealth program can play a pivotal role in providing emotional support for HF patients. Considering the close relationship between emotional health and hard clinical outcomes in HF [44,45], ICT-based telehealth program has the potential to improve the prognosis of HF patients. The improvement in physical subscores was also numerically greater in patients with good adherence without reaching statistical significance. Considering that patients with good adherence to ICT-based telehealth program showed reduced uNa, a surrogate marker of better control of sodium intake, it seems plausible that this program can also be helpful in improving the physical aspects of quality of life in HF patients. A future study with a large sample size is needed to validate these findings.

Importance of Adherence in Self-Care Management

Previous studies demonstrated that self-care management using telemonitoring systems did not significantly improve hard endpoints in HF patients, such as all-cause death or HF hospitalization [46,47]. The low rate of adherence to self-care management programs has been suggested as one of possible reasons for these suboptimal results. In the tele-HF trial, only 55% of study patients utilized the telemonitoring system at least 3 times per week during 6 months and, furthermore, 14% of patients never used this system during the study period [46]. In our study, 23 of 27 patients (85.2%) utilized ICT-based telehealth program at least 3 times per week during 12 weeks, and there were no patients who did not use the program. Furthermore, 10 of 27 patients (37.0%) used the ICT-based telehealth program ≥100 times during the 12-week period, indicating the use of this program more than once daily. These patients with good adherence to the program showed the improvement in sodium intake and quality of life. This improved adherence of patients may be partially due to the enhanced integration of ICT-based telehealth program into their daily lives, with the easier use of such systems.

User Experience With Voice Recognition Component of the Technology

The user experience was positive with ≥85% patients indicating satisfaction or neutral response with the voice recognition system provided. This favorable user experience might have been achieved by the acceptable success rate of voice recognition in our study, considering that 95% accuracy is nearly on par with that of human speech recognition. In this study, however, the accuracy of voice recognition was affected by adverse conditions, including the use of nonstandard speech, voice alteration during walking, and the presence of substantial background noise. Specifically, the success rate of voice recognition was 91% and 86% for the patients who used nonstandard language and for those who were walking during the process of voice recognition, respectively. Furthermore, the success rate was 87% in settings with increased background noise level, such as when using public transportation. Hence, the user experience can be further optimized by improving the accuracy of voice recognition under these unfavorable circumstances.

Implications of Using ICT-Based Telehealth Program With Voice Recognition Technology on Study Endpoints

HF patients performing and maintaining self-care management behaviors, such as control of sodium intake (uNa), is one practical implication of this technology from a health care perspective. Although future studies with larger number of HF patients are warranted, this technology also has the potential to improve health-related quality of life (MLHFQ scores). Current implications for improving functional capacity (6-min walk test) or HF severity (NT-proBNP) are limited. Further studies are also needed to determine whether the use of this technology has workforce implications in the management of HF patients.

Study Limitations

This study has several limitations that should be considered. First, our sample size was small and the duration of intervention was short. This limitation made it difficult to show the benefit of the ICT-based telehealth program for clinical outcomes in HF patients, especially given the chronic and progressive nature of HF. However, in the case of pilot studies, sample sizes of 10 to 30 are generally considered large enough to test the null hypothesis and small enough to ignore weak treatment effects [47]. Furthermore, even though our study population largely comprised patients with moderate to severe systolic dysfunction (LVEF <40%) whose mortality is substantial (25.6%-41.7%) [48], there was no mortality and only 2 patients withdrew because of hospitalization for acute decompensated HF at the early phase of the study. Notably, the most likely cause of HF exacerbation in these patients was viral infection, not a high-salt diet or poor adherence to medication. In addition, there were no rehospitalizations or emergency department visits for HF after 4 weeks of intervention in our study participants, approximately 75% of whom had a history of prior HF hospitalizations, suggesting that clinical outcomes can be improved by applying this strategy to HF patients. Considering that a prior history of HF hospitalizations has been reported as a risk factor for repeat hospitalizations in HF patients [49,50], our findings suggest that clinical outcomes can be improved by the appropriate application of ICT-based telehealth program in HF patients, although larger studies with longer follow-up are definitely warranted. Second, the definition of good adherence was arbitrary. We assumed that the total frequency of program use of more than 100 times during the follow-up of 12 weeks could be regarded as accessing the system at least once per day. Third, we did not have data on baseline education level, including literacy, health care literacy, digital literacy, and numeracy skills of the study population, which can be critical components for successful technology uptake [51]. Fourth, although spot uNa provides an approximate estimate of sodium excretion, it has limited usefulness as an indicator of long-term dietary behavior change because of its short-term relevance. Thus, a future study using a more accurate surrogate marker, such as 24-hour urine sodium excretion, is warranted to more clearly discern the beneficial effect of ICT-based telehealth program on sodium intake. Finally, because of the lack of an appropriate control arm, study findings should be interpreted with caution. A prospective randomized study comparing an ICT-based telehealth program with conventional HF care is necessary to define the true clinical benefit from this therapeutic strategy in HF patients.

Conclusions

Short-term application of ICT-based telehealth program in HF demonstrated the potential to improve control of sodium intake and quality of life in chronic HF patients with good adherence to this program. Our results suggest that an ICT-based telehealth program can be an effective strategy to achieve better patient care and clinical outcomes in HF.

Acknowledgments

This study was partly supported by a grant from the Ministry of Science, ICT and Future Planning, Republic of Korea (F1101-13-1010).

Conflicts of Interest

Jong-Hyuk Choi and Young-Joon Lee hold a patent of invention for voice recognition technology applied to HF self-care management (Korea patent number 10-1141605). They are also employees and shareholders of AIMMED Co Ltd, which provides this ICT service. Other authors report no conflicts of interest to disclose.

Multimedia Appendix 1

The ICT-based telehealth program website.

[PDF File (Adobe PDF File), 182KB - mhealth_v5i10e127_app1.pdf]

Multimedia Appendix 2

Comparison of characteristics of ICT-based telehealth program with voice recognition technology and previous programs.

[PDF File (Adobe PDF File), 27KB - mhealth_v5i10e127_app2.pdf]

Multimedia Appendix 3

Patient satisfaction with the voice recognition component of the technology.

[PDF File (Adobe PDF File), 261KB - mhealth_v5i10e127_app3.pdf]

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation 2015 Jan 27;131(4):e29-322. [doi: 10.1161/CIR.0000000000000152] [Medline: 25520374]
2. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol 1993 Oct;22(4 Suppl A):6A-13A. [Medline: 8376698]
3. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013 Oct 15;62(16):e147-e239 [FREE Full text] [doi: 10.1016/j.jacc.2013.05.019] [Medline: 23747642]
4. Setoguchi S, Stevenson LW, Schneeweiss S. Repeated hospitalizations predict mortality in the community population with heart failure. Am Heart J 2007 Aug;154(2):260-266. [doi: 10.1016/j.ahj.2007.01.041] [Medline: 17643574]
5. Goodlin SJ. Palliative care in congestive heart failure. J Am Coll Cardiol 2009 Jul 28;54(5):386-396 [FREE Full text] [doi: 10.1016/j.jacc.2009.02.078] [Medline: 19628112]
6. Kociol RD, Peterson ED, Hammill BG, Flynn KE, Heidenreich PA, Piña IL, et al. National survey of hospital strategies to reduce heart failure readmissions: findings from the get With the guidelines-heart failure registry. Circ Heart Fail 2012 Nov;5(6):680-687 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.112.967406] [Medline: 22933525]
7. Gheorghiade M, Vaduganathan M, Fonarow GC, Bonow RO. Rehospitalization for heart failure: problems and perspectives. J Am Coll Cardiol 2013 Jan 29;61(4):391-403 [FREE Full text] [doi: 10.1016/j.jacc.2012.09.038] [Medline: 23219302]
8. Currie K, Strachan PH, Spaling M, Harkness K, Barber D, Clark AM. The importance of interactions between patients and healthcare professionals for heart failure self-care: a systematic review of qualitative research into patient perspectives. Eur J Cardiovasc Nurs 2015 Dec;14(6):525-535. [doi: 10.1177/1474515114547648] [Medline: 25139468]
9. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, et al. 2016 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European society of cardiology (ESC) developed with the special contribution of the heart failure association (HFA) of the ESC. Eur Heart J 2016 Jul 14;37(27):2129-2200. [doi: 10.1093/eurheartj/ehw128] [Medline: 27206819]
10. Lainscak M, Blue L, Clark AL, Dahlström U, Dickstein K, Ekman I, et al. Self-care management of heart failure: practical recommendations from the patient care committee of the heart failure association of the European society of cardiology. Eur J Heart Fail 2011 Feb;13(2):115-126. [doi: 10.1093/eurjhf/hfq219] [Medline: 21148593]
11. Mead H, Andres E, Ramos C, Siegel B, Regenstein M. Barriers to effective self-management in cardiac patients: the patient's experience. Patient Educ Couns 2010 Apr;79(1):69-76. [doi: 10.1016/j.pec.2009.08.003] [Medline: 19748205]
12. Inglis SC, Clark RA, McAlister FA, Stewart S, Cleland JGF. Which components of heart failure programmes are effective? a systematic review and meta-analysis of the outcomes of structured telephone support or telemonitoring as the primary component of chronic heart failure management in 8323 patients: abridged cochrane review. Eur J Heart Fail 2011 Sep;13(9):1028-1040. [doi: 10.1093/eurjhf/hfr039] [Medline: 21733889]
13. Jerant AF, Azari R, Nesbitt TS. Reducing the cost of frequent hospital admissions for congestive heart failure: a randomized trial of a home telecare intervention. Med Care 2001 Nov;39(11):1234-1245. [Medline: 11606877]
14. Adams WG, Fuhlbrigge AL, Miller CW, Panek CG, Gi Y, Loane KC, et al. TLC-Asthma: an integrated information system for patient-centered monitoring, case management, and point-of-care decision support. AMIA Annu Symp Proc 2003:1-5 [FREE Full text] [Medline: 14728122]
15. Kim KM, Park KS, Lee HJ, Lee YH, Bae JS, Lee YJ, et al. Efficacy of a new medical information system, ubiquitous healthcare service with voice inception technique in elderly diabetic patients. Sci Rep 2015 Dec 11;5:18214 [FREE Full text] [doi: 10.1038/srep18214] [Medline: 26658492]

16. Glanz K, Beck AD, Bundy L, Primo S, Lynn MJ, Cleveland J, et al. Impact of a health communication intervention to improve glaucoma treatment adherence. Results of the interactive study to increase glaucoma adherence to treatment trial. Arch Ophthalmol 2012 Oct;130(10):1252-1258 [FREE Full text] [doi: 10.1001/archophthalmol.2012.1607] [Medline: 22688429]
17. Alla F, Briançon S, Guillemin F, Juillière Y, Mertès P, Villemot J, et al. Self-rating of quality of life provides additional prognostic information in heart failure. Insights into the EPICAL study. Eur J Heart Fail 2002 Jun;4(3):337-343 [FREE Full text] [Medline: 12034160]
18. Ni H, Toy W, Burgess D, Wise K, Nauman DJ, Crispell K, et al. Comparative responsiveness of short-form 12 and Minnesota living with heart failure questionnaire in patients with heart failure. J Card Fail 2000 Jun;6(2):83-91. [Medline: 10908081]
19. Rector TS, Cohn JN. Assessment of patient outcome with the Minnesota living with heart failure questionnaire: reliability and validity during a randomized, double-blind, placebo-controlled trial of pimobendan. Pimobendan multicenter research group. Am Heart J 1992 Oct;124(4):1017-1025. [Medline: 1529875]
20. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the task force for the diagnosis and treatment of acute and chronic heart failure 2012 of the European society of cardiology. Developed in collaboration with the heart failure association (HFA) of the ESC. Eur Heart J 2012 Jul;33(14):1787-1847 [FREE Full text] [doi: 10.1093/eurheartj/ehs104] [Medline: 22611136]
21. Keell SD, Chambers JS, Francis DP, Edwards DF, Stables RH. Shuttle-walk test to assess chronic heart failure. Lancet 1998 Aug 29;352(9129):705. [Medline: 9728991]
22. Rasekaba T, Lee AL, Naughton MT, Williams TJ, Holland AE. The six-minute walk test: a useful metric for the cardiopulmonary patient. Intern Med J 2009 Aug;39(8):495-501. [doi: 10.1111/j.1445-5994.2008.01880.x] [Medline: 19732197]
23. Skorecki KL, Brenner BM. Body fluid homeostasis in man. A contemporary overview. Am J Med 1981 Jan;70(1):77-88. [Medline: 7006395]
24. Verbrugge FH, Dupont M, Steels P, Grieten L, Swennen Q, Tang WH, et al. The kidney in congestive heart failure: 'are natriuresis, sodium, and diuretics really the good, the bad and the ugly?'. Eur J Heart Fail 2014 Feb;16(2):133-142 [FREE Full text] [doi: 10.1002/ejhf.35] [Medline: 24464967]
25. Luft FC, Fineberg NS, Sloan RS. Estimating dietary sodium intake in individuals receiving a randomly fluctuating intake. Hypertension 1982;4(6):805-808 [FREE Full text] [Medline: 7141607]
26. Kang SS, Kang EH, Kim SO, Lee MS, Hong CD, Kim SB. Use of mean spot urine sodium concentrations to estimate daily sodium intake in patients with chronic kidney disease. Nutrition 2012 Mar;28(3):256-261. [doi: 10.1016/j.nut.2011.06.006] [Medline: 21996048]
27. Ogura M, Kimura A, Takane K, Nakao M, Hamaguchi A, Terawaki H, et al. Estimation of salt intake from spot urine samples in patients with chronic kidney disease. BMC Nephrol 2012 Jun 08;13:36 [FREE Full text] [doi: 10.1186/1471-2369-13-36] [Medline: 22682402]
28. McLean R, Williams S, Mann J. Monitoring population sodium intake using spot urine samples: validation in a New Zealand population. J Hum Hypertens 2014 Nov;28(11):657-662. [doi: 10.1038/jhh.2014.10] [Medline: 24573130]
29. Brown IJ, Dyer AR, Chan Q, Cogswell ME, Ueshima H, Stamler J, et al. Estimating 24-hour urinary sodium excretion from casual urinary sodium concentrations in Western populations: the INTERSALT study. Am J Epidemiol 2013 Jun 1;177(11):1180-1192 [FREE Full text] [doi: 10.1093/aje/kwt066] [Medline: 23673246]
30. Averina VA, Othmer HG, Fink GD, Osborn JW. A new conceptual paradigm for the haemodynamics of salt-sensitive hypertension: a mathematical modelling approach. J Physiol 2012 Dec 01;590(23):5975-5992 [FREE Full text] [doi: 10.1113/jphysiol.2012.228619] [Medline: 22890716]
31. Brown IJ, Tzoulaki I, Candeias V, Elliott P. Salt intakes around the world: implications for public health. Int J Epidemiol 2009 Jun;38(3):791-813 [FREE Full text] [doi: 10.1093/ije/dyp139] [Medline: 19351697]
32. Mann SJ, Gerber LM. Estimation of 24-hour sodium excretion from spot urine samples. J Clin Hypertens (Greenwich) 2010 Mar;12(3):174-180 [FREE Full text] [doi: 10.1111/j.1751-7176.2009.00241.x] [Medline: 20433530]
33. Kawano Y, Tsuchihashi T, Matsuura H, Ando K, Fujita T, Ueshima H, Working Group for Dietary Salt Reduction of the Japanese Society of Hypertension. Report of the working group for dietary salt reduction of the Japanese society of hypertension: (2) assessment of salt intake in the management of hypertension. Hypertens Res 2007 Oct;30(10):887-893. [doi: 10.1291/hypres.30.887] [Medline: 18049019]
34. Verbrugge FH, Nijst P, Dupont M, Penders J, Tang WHW, Mullens W. Urinary composition during decongestive treatment in heart failure with reduced ejection fraction. Circ Heart Fail 2014 Sep;7(5):766-772. [doi: 10.1161/CIRCHEARTFAILURE.114.001377] [Medline: 25037309]
35. Ferreira JP, Girerd N, Medeiros PB, Santos M, Carvalho HC, Bettencourt P, et al. Spot urine sodium excretion as prognostic marker in acutely decompensated heart failure: the spironolactone effect. Clin Res Cardiol 2016 Jun;105(6):489-507. [doi: 10.1007/s00392-015-0945-x] [Medline: 26615605]
36. Heart Failure Society Of America. Nonpharmacologic management and health care maintenance in patients with chronic heart failure. J Card Fail 2006 Feb;12(1):e29-e37. [doi: 10.1016/j.cardfail.2005.11.011] [Medline: 16500569]

37. Piotrowicz E, Baranowski R, Bilinska M, Stepnowska M, Piotrowska M, Wójcik A, et al. A new model of home-based telemonitored cardiac rehabilitation in patients with heart failure: effectiveness, quality of life, and adherence. Eur J Heart Fail 2010 Feb;12(2):164-171. [doi: 10.1093/eurjhf/hfp181] [Medline: 20042423]
38. Leidy N, Rentz A, Zyczynski T. Evaluating Health-related quality-of-life outcomes in patients with congestive heart failure. A review of recent randomised controlled trials. Pharmacoeconomics 1999;15:19-46. [Medline: 10345157]
39. Johnson G, Carson P, Francis GS, Cohn JN. Influence of prerandomization (baseline) variables on mortality and on the reduction of mortality by enalapril. Veterans Affairs Cooperative Study on Vasodilator Therapy of Heart Failure (V-HeFT II). V-HeFT VA Cooperative Studies Group. Circulation 1993 Jun;87(6 Suppl):VI32-VI39. [Medline: 8500237]
40. Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, et al. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. Heart Fail Rev 2014 May;19(3):359-367. [doi: 10.1007/s10741-013-9394-7] [Medline: 23681849]
41. Rector TS, Tschumperlin LK, Kubo SH, Bank AJ, Francis GS, McDonald KM, et al. Use of the Living With Heart Failure questionnaire to ascertain patients' perspectives on improvement in quality of life versus risk of drug-induced death. J Card Fail 1995 Jun;1(3):201-206. [Medline: 9420652]
42. Hoekstra T, Jaarsma T, van VD, Hillege HL, Sanderman R, Lesman-Leegte I. Quality of life and survival in patients with heart failure. Eur J Heart Fail 2013 Jan;15(1):94-102 [FREE Full text] [doi: 10.1093/eurjhf/hfs148] [Medline: 22989869]
43. Wu RC, Thorpe K, Ross H, Micevski V, Marquez C, Straus SE. Comparing administration of questionnaires via the internet to pen-and-paper in patients with heart failure: randomized controlled trial. J Med Internet Res 2009;11(1):e3 [FREE Full text] [doi: 10.2196/jmir.1106] [Medline: 19275979]
44. Mentz RJ, Babyak MA, Bittner V, Fleg JL, Keteyian SJ, Swank AM, HF-ACTION Investigators. Prognostic significance of depression in blacks with heart failure: insights from heart failure: a controlled trial investigating outcomes of exercise training. Circ Heart Fail 2015 May;8(3):497-503 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.114.001995] [Medline: 25901047]
45. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
46. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):e34 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
47. Julious S. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics 2005;4:287-291. [doi: 10.1002/pst.185]
48. Curtis JP, Sokol SI, Wang Y, Rathore SS, Ko DT, Jadbabaie F, et al. The association of left ventricular ejection fraction, mortality, and cause of death in stable outpatients with heart failure. J Am Coll Cardiol 2003 Aug 20;42(4):736-742 [FREE Full text] [Medline: 12932612]
49. Curtis LH, Greiner MA, Hammill BG, Kramer JM, Whellan DJ, Schulman KA, et al. Early and long-term outcomes of heart failure in elderly persons, 2001-2005. Arch Intern Med 2008 Dec 08;168(22):2481-2488 [FREE Full text] [doi: 10.1001/archinte.168.22.2481] [Medline: 19064833]
50. Dunlay SM, Redfield MM, Weston SA, Therneau TM, Hall LK, Shah ND, et al. Hospitalizations after heart failure diagnosis a community perspective. J Am Coll Cardiol 2009 Oct 27;54(18):1695-1702 [FREE Full text] [doi: 10.1016/j.jacc.2009.08.019] [Medline: 19850209]
51. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]

Abbreviations

ACEI: angiotensin-converting enzyme inhibitor
ACS: automatic call system
ARB: angiotensin receptor blocker
ARS: automatic response system
BP: blood pressure
CO2: carbon dioxide
EF: ejection fraction
GFR: glomerular filtration rate
HF: heart failure
ICT: information communication technology
LV: left ventricular
MLHFQ: Minnesota Living With Heart Failure Questionnaire
NT-proBNP: N-terminal prohormone of brain natriuretic peptide
NYHA: New York Heart Association

SE: standard error
SMS: short message service
TTS: text to speech
uNa: urine sodium concentration
WBC: white blood cell

Edited by CL Parra-Calderón; submitted 28.11.16; peer-reviewed by H Ross, H Al-Samarraie; comments to author 07.02.17; revised version received 30.06.17; accepted 25.07.17; published 02.10.17.

Please cite as:
Lee H, Park JB, Choi SW, Yoon YE, Park HE, Lee SE, Lee SP, Kim HK, Cho HJ, Choi SY, Lee HY, Choi J, Lee YJ, Kim YJ, Cho GY, Choi J, Sohn DW
Impact of a Telehealth Program With Voice Recognition Technology in Patients With Chronic Heart Failure: Feasibility Study
JMIR Mhealth Uhealth 2017;5(10):e127
URL: https://mhealth.jmir.org/2017/10/e127/
doi:10.2196/mhealth.7058
PMID:28970189

Original Paper

Fall Detection in Individuals With Lower Limb Amputations Using Mobile Phones: Machine Learning Enhances Robustness for Real-World Applications

Nicholas Shawen[1,2*], MS; Luca Lonini[1,2,3*], PhD; Chaithanya Krishna Mummidisetty[1], MS; Ilona Shparii[1,2,4], MS; Mark V Albert[1,2,3,4], PhD; Konrad Kording[5,6], PhD; Arun Jayaraman[1,2,3,7], PT, PhD

[1]Max Nader Lab for Rehabilitation Technologies and Outcomes Research, Shirley Ryan AbilityLab, Chicago, IL, United States

[2]Center for Bionic Medicine, Shirley Ryan AbilityLab, Chicago, IL, United States

[3]Department of Physical Medicine and Rehabilitation, Northwestern University, Chicago, IL, United States

[4]Department of Computer Science, Loyola University Chicago, Chicago, IL, United States

[5]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States

[6]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, United States

[7]Department of Physical Therapy and Human Movement Sciences, Northwestern University, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Luca Lonini, PhD
Max Nader Lab for Rehabilitation Technologies and Outcomes Research
Shirley Ryan AbilityLab
355 E Erie St
Suite #11-1101
Chicago, IL, 60611
United States
Phone: 1 312 238 1619
Email: llonini@ricres.org

Related Article:

This is a corrected version. See correction statement: http://mhealth.jmir.org/2017/11/e167/

Abstract

Background: Automatically detecting falls with mobile phones provides an opportunity for rapid response to injuries and better knowledge of what precipitated the fall and its consequences. This is beneficial for populations that are prone to falling, such as people with lower limb amputations. Prior studies have focused on fall detection in able-bodied individuals using data from a laboratory setting. Such approaches may provide a limited ability to detect falls in amputees and in real-world scenarios.

Objective: The aim was to develop a classifier that uses data from able-bodied individuals to detect falls in individuals with a lower limb amputation, while they freely carry the mobile phone in different locations and during free-living.

Methods: We obtained 861 simulated indoor and outdoor falls from 10 young control (non-amputee) individuals and 6 individuals with a lower limb amputation. In addition, we recorded a broad database of activities of daily living, including data from three participants' free-living routines. Sensor readings (accelerometer and gyroscope) from a mobile phone were recorded as participants freely carried it in three common locations—on the waist, in a pocket, and in the hand. A set of 40 features were computed from the sensors data and four classifiers were trained and combined through stacking to detect falls. We compared the performance of two population-specific models, trained and tested on either able-bodied or amputee participants, with that of a model trained on able-bodied participants and tested on amputees. A simple threshold-based classifier was used to benchmark our machine-learning classifier.

Results: The accuracy of fall detection in amputees for a model trained on control individuals (sensitivity: mean 0.989, 1.96*standard error of the mean [SEM] 0.017; specificity: mean 0.968, SEM 0.025) was not statistically different (P=.69) from that of a model trained on the amputee population (sensitivity: mean 0.984, SEM 0.016; specificity: mean 0.965, SEM 0.022).

Detection of falls in control individuals yielded similar results (sensitivity: mean 0.979, SEM 0.022; specificity: mean 0.991, SEM 0.012). A mean 2.2 (SD 1.7) false alarms per day were obtained when evaluating the model (vs mean 122.1, SD 166.1 based on thresholds) on data recorded as participants carried the phone during their daily routine for two or more days. Machine-learning classifiers outperformed the threshold-based one (P<.001).

Conclusions: A mobile phone-based fall detection model can use data from non-amputee individuals to detect falls in individuals walking with a prosthesis. We successfully detected falls when the mobile phone was carried across multiple locations and without a predetermined orientation. Furthermore, the number of false alarms yielded by the model over a longer period of time was reasonably low. This moves the application of mobile phone-based fall detection systems closer to a real-world use case scenario.

(JMIR Mhealth Uhealth 2017;5(10):e151) doi:10.2196/mhealth.8201

KEYWORDS

fall detection; lower limb amputation; mobile phones; machine learning

Introduction

Falls are a common occurrence in the elderly and in people with lower limb amputations. In the elderly, they are the primary cause of injury-related deaths [1], and for people older than 75 years, the estimated percentage who fall is more than 30% per year [2]. Individuals with an amputation, especially elderly with an amputation due to vascular disease, are at a similar or higher risk for falls, with studies reporting that more than 50% of individuals with a unilateral lower limb amputation had fallen in the previous 12 months [3,4]. As such, detecting, understanding, and reacting to falls are of very high importance.

Decreasing response time after a fall and improving fall prevention strategies can dramatically enhance the quality of life for people with lower limb amputations, as well as decrease health care costs. Getting help following an immobilizing fall improves the chance of survival by approximately 50% and increases the likelihood of a return to independent living [5]. Therefore, detecting falls and understanding the environmental circumstances that led to it is crucial for both timely assistance and evaluation of prevention strategies.

Data from real-world fall events are essential for such analyses. However, capturing data from real-world falls is difficult without long-term continuous monitoring [6]. Mobile phones can provide an inexpensive way to detect and measure falls over long time periods [7-10]. All modern mobile phones are equipped with multiple sensors—most notably accelerometers, gyroscopes, barometer, and Global Positioning System (GPS)—which generate a wide range of information regarding the user's movements and location. Furthermore, mobile phones include memory, computing, and transmission capabilities, which makes them a convenient platform to process the sensors' data and detect falls [11]. Mobile phones then promise a relatively straightforward generation of large datasets because they can unobtrusively record all the time.

Sensor-based fall-detection algorithms have shown encouraging results in previous studies using young, unimpaired participants performing simulated falls [12,13]. Many of these studies employed threshold-based algorithms, such that a fall is detected if one or more statistical measures (features) computed from the acceleration exceeded a predefined threshold [6,14], whereas some approaches employed machine-learning algorithms, either supervised [15-18] or unsupervised [19]. Such studies showed that simulated falls can be successfully distinguished from daily activities with high accuracy under laboratory-controlled conditions.

Traditionally, studies have used unimpaired individuals for training and testing fall-detection systems. However, a large difference exists between movement patterns of individuals with and without lower limb amputations [20]. Furthermore, most studies have fixed the phone to a specific orientation and location on the body, although phones are carried in multiple locations during everyday use. Common movements, such as taking the phone out of a pocket, could cause large accelerations that may be confused with falls. Therefore, we do not know whether these factors affect the accuracy of a fall-detection algorithm when used with individuals with an amputation.

In this study, we developed a fall-detection classifier that is robust to the previously mentioned sources of error (population, location of the phone, environment) and successfully detect falls in both control (non-amputee) and amputee populations. We collected data from simulated falls and activities from both control volunteers and individuals with transfemoral amputations (TFAs) or above-the-knee amputations in both a laboratory and outdoor environment. To account for the influence of phone location, we collected simulated falls and activities data with the mobile phone carried in three different common locations: in a pouch at the waist, in a pocket, and in the participant's hand.

Methods

Study Design

Data representative of typical activities of daily living and falls was collected from both control participants and participants with a unilateral TFA using a prosthesis. Participants with a TFA were included in the study if they had a unilateral amputation of the lower limb, above or below the knee, within at least 6 months, and if they used either a mechanical or microprocessor-controlled prosthesis on a daily basis. The study was approved by Northwestern University Institutional review board. Written informed consent was obtained from all participants.

Data were collected from the accelerometer and gyroscope sensor of a mobile phone (Samsung Galaxy S4) using the Purple Robot app [21] running on Android 4.4.4. Purple Robot was developed as a research platform for collecting data through

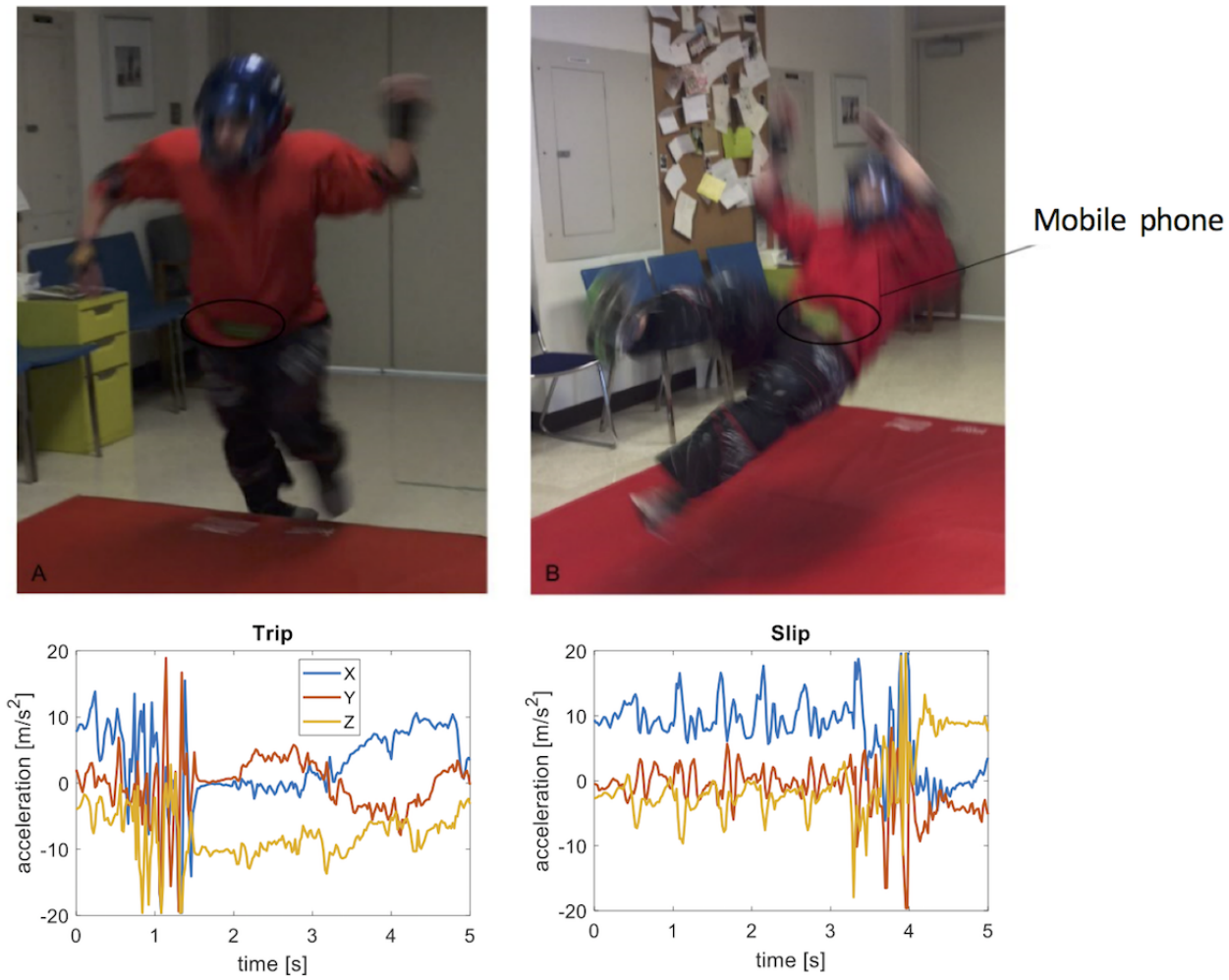
hardware sensors on an Android mobile phone. Data from the selected sensors are compiled and transmitted to a remote server for storage and future analysis via Wi-Fi or cellular data connection. The app also allows for some data processing to occur on the phone.

The data sampling rate was approximately 50 Hz and could vary depending on the phone central processing unit usage. Participants carried the phone in three different locations during the data collection: in a pouch worn on the waist, in a pants pocket, or in their hand. Participants were asked to carry the phone in these locations as they would carry their own phone during daily use; as a result, there was some variation in the precise placement of the device (eg, the pocket location included both front and back pockets based on individual preference). During data collection, a researcher annotated the start and end time of each activity or fall with a second mobile phone running Purple Robot.

All participants performed four types of simulated falls: forward (trip), backward (slip), left, and right. Non-amputee participants performed both indoor and outdoor falls, whereas participants with amputation only fell indoors. During all simulated falls, participants fell onto a padded mat (indoor) or grass (outdoor), and wore several layers of padding and guards over the wrists, elbows, knees, and shins to prevent injury. During simulated trips, participants were asked to walk toward the mat and then stumble on the edge as if tripping, then falling forward onto the mat. For slips, participants were instructed to slide one foot out from under them as if slipping and fall backward onto the mat (Figure 1). For lateral falls (left and right), participants were asked to close their eyes and a researcher provided a push to cause the participant to lose their balance and fall onto the mat. The TFA participants were instructed to use whatever protective strategies they might employ in a real fall event, such as turning to avoid landing on their prosthesis. Participants performed each fall type three times for each phone location, for a total of 36 simulated falls.

After completing the falls, participants performed a series of daily activities, again with the phone varied between the three different locations (see Multimedia Appendix 1 for a full list of activities). The different activities performed were sitting, standing, walking, stairs ascent/descent, and lying. During still activities with the phone in hand, participants were asked to use the phone as they usually would (eg, browsing the Internet, checking text messages). In total, approximately 17 hours of labeled data from falls and activities were obtained. In addition, three participants with TFA carried the phone with them for a period between two and seven days, so as to quantify the number of false alarms per day generated by the fall-detection model. No specific instructions were given to the participants, beyond that of carrying the phone in either their pocket or in a pouch around the waist for the majority of their daily routine. All three participants chose to carry the phone in a pants pocket. Over the entire recording period, a mean 213 (SD 182) clips per day (total 2251) exceeded the 2 *g* threshold and were subsequently analyzed.

Figure 1. Experimental setup for capturing falls data and detecting falls. A non-amputee volunteer performed a series of falls, including trips (left) and slips (right). A phone was carried in a pouch secured around the waist with a strap belt and in a pants pocket or in hand (not shown). The graphs show example data captured by the phone accelerometer during the two types of falls over a 5 second window.



Data Preprocessing and Feature Extraction

Data collected by Purple Robot was transmitted to an external server and downloaded for analysis using Matlab 2016b (MathWorks) and Python 2.7. All data were organized into 5-second clips of data for feature extraction. For each fall event, a 10-second long data clip centered on the peak acceleration magnitude (ie, the impact) was extracted; from these clips, ten 5-second long windows were extracted based on a uniform random distribution, such that the beginning of the fall in each 5-second window (clip) can occur with uniform probability in the interval (0 s, 3 s). This was done to provide variety in the location of the falls within a 5-second window.

Activities data were generated by taking all data from the beginning to the end of the activities protocol and dividing it into nonoverlapping 5-second windows. Thus, our activity data included postural transitions (eg, sit-to-stand) and phone transitions (eg, pocket-to-hand) that may be confounded with falls by a classifier. We then selected activity clips whose total acceleration ($x^2+y^2+z^2$) was higher than 2 *g*, which corresponded to the first percentile of the acceleration distribution within the falls (Figure 2). Therefore, only activity clips that included high accelerations and could thus resemble a fall were included in the dataset. In total, 6637 clips (6337 falls, 300 non-falls) of data were obtained from the non-amputee control group and 1815 clips (1537 falls, 278 non-falls) from the TFA group.

After organizing data into the 5-second windows, the accelerometer and gyroscope signals were interpolated to 50 Hz with a cubic polynomial. A total of 40 features were then computed on each axis (x, y, z) and on the resultant vectors ($x^2+y^2+z^2$) for both the accelerometer and gyroscope signals (Table 1). In training and testing our fall-detection models, all feature vectors derived from activities were labeled as "non-falls" and those derived from simulated falls as "falls."

Model Training and Evaluation

We combined the predictions of four different classifiers through stacking [22]: random forest (100 trees), support vector machine [23] (linear kernel, C=1), gradient boosting (100 trees, maximum depth=2), and extreme gradient boosting (XGBoost [24], 150 trees, maximum depth=2, learning rate=0.5, feature subsampling rate=0.6). Each classifier (C1,...,C4) predicts the probability $P_i(\text{fall}|x)$ that the current clip *x* corresponds to a fall, and all the probabilities are combined into a feature vector:

$$x_{meta}=[P_1(\text{fall}|x),...,P_4(\text{fall}|x),\sigma(P_i)]$$

where $i=1,...,4$, and these probabilities and their standard deviation (σ) are used as input features to a meta-level classifier

(logistic regression [25,26]), which learns to combine the individual predictions and outputs the final probability of the clip *x* being a fall (Figure 3).

Classifier performance was evaluated either by leave-one-subject-out cross-validation (LOSOCV) or with an external validation set (see next section). Hyperparameters of the base-level classifiers were tuned using a cross-validated grid search. The primary measures chosen to summarize model performance were sensitivity, specificity, and area under the curve (AUC). Specificity and sensitivity values are reported from the optimal point from the receiver-operator characteristic (ROC) curve, or the point with the greatest sum of sensitivity and specificity. LOSOCV was used for determining the threshold when evaluating the control-to-control model.

As a comparison, we evaluated the performance obtained when using a threshold on a single feature, the maximum acceleration magnitude. This was chosen as a high-performing representative of threshold-based methods because the acceleration magnitude is one the strongest single predictors of a fall and is frequently used [27]. An ROC curve was obtained by measuring the sensitivity and specificity as the threshold on maximum acceleration magnitude was changed, from which we obtained values for AUC, sensitivity, and specificity for this single-feature approach.

Table 1. Features computed on each 5-second clip of sensor data (accelerometer and gyroscope) on either the vector resultant or on each axis (x, y, z).

Feature name	Number of features
Mean	1
Median	1
Standard deviation	1
Skewness	1
Kurtosis	1
IQR and derivative of IQR	2
Minimum and derivative of minimum	2
Maximum and derivative of maximum	2
Maximum, minimum, and IQR on each axis (x, y, z)	9
Total per sensor	20

Figure 2. The distribution of maximum acceleration values for falls and non-fall data clips. Values for non-amputee (left) and individuals with TFA (right). Only activity (non-fall) clips with an acceleration greater than 2g were used in the analysis. Peak accelerations from participants' daily routine (home trial) are also shown for the TFA group. Boxes indicate interquartile (IQR) range; midlines and whiskers represent median and 1.5 IQR, respectively. Individual points denote outliers.

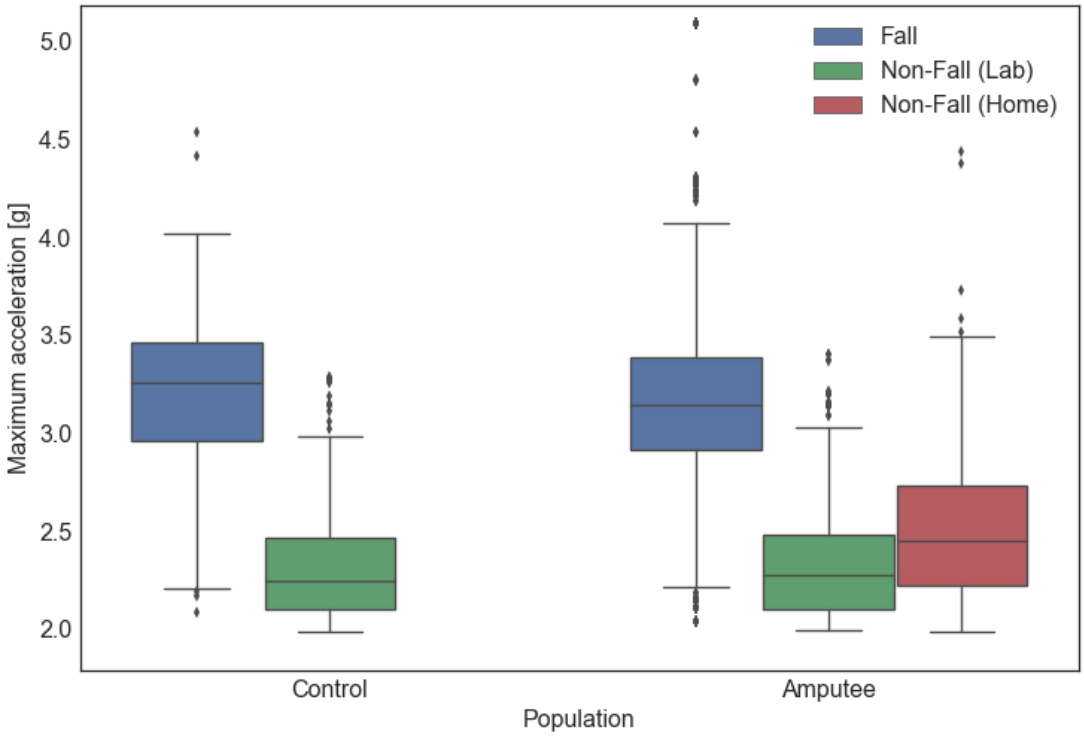
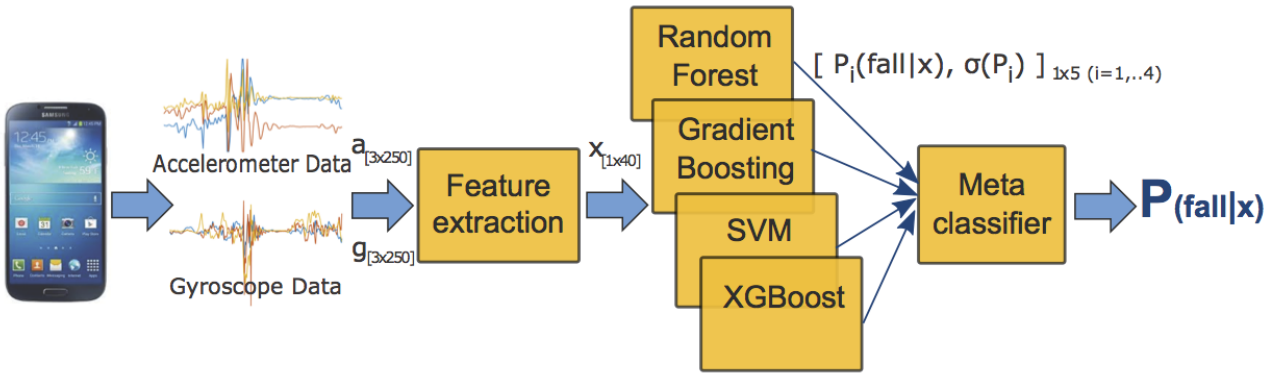


Figure 3. Five-second data clips are recorded from the mobile phone sensors (accelerometer and gyroscope), with each clip yielding a matrix of dimension (3 channels × 250 samples) per sensor. A set of 40 features were calculated from a data clip, and the resulting feature vector x was input to four different classifiers, which were combined through stacking to output the probability of the clip being a fall (see text for details).



Effect of Training Population and Phone Location

To determine whether data collected from control (non-amputee) individuals will be effective to detect falls in amputees, we trained and tested models under three training conditions: a model trained and evaluated on control participants using LOSOCV (control to control), a model trained using data from all control participants and evaluated on data from the TFA group (control to amputee), and a model trained and evaluated on TFA participants using LOSOCV (amputee to amputee). Each model was trained and evaluated using data collected in all three phone locations. Furthermore, we assessed the performance of the control-to-amputee model at detecting falls from each individual location (waist, pocket, or hand). Each model was also compared to its corresponding threshold-based version, which represented the baseline performance.

Measuring Performance at Home

Our model is intended for real-time use on a mobile phone; therefore, we also collected data from TFA participants while they carried the phone in a pocket for at least 2 days. No participant fell during this data collection, so the data represents only non-fall events. This dataset was filtered as before, generating a total of 2467 data clips above the 2 *g* threshold. Resampling and feature extraction were performed as previously to prepare the data for model evaluation. The preceding analysis procedure was repeated, using data collected at home to represent non-falls, rather than daily activities performed in the laboratory.

Results

A total of 7 amputees (mean 47.4, SD 12.0 years) and 10 control non-amputee participants (mean 24.2, SD 2.2 years) took part to the study. One amputee participant withdrew before performing the outdoor falls protocol and their data was therefore excluded from the analysis. Also, two amputee participants could not complete the entire set of falls because of fatigue. Table 2 describes participant demographics for both groups.

Table 2. Demographic information of participants.

Subject ID	Age (years)	Gender	Height (ft/in)	Weight (lbs)	Amputation side	Amputation reason	Type of prosthesis
Amputee participants							
AF004	58	Male	5'10“	203	Left	Trauma	Mechanical
AF005	51	Female	5'5”	160	Right	Cancer	Microprocessor
AF006	24	Male	5'10“	205	Right	Cancer	Microprocessor
AF007	54	Male	6'1”	240	Left	Trauma	Hydraulic
AF008	37	Female	5'3“	102	Right	Congenital	Mechanical
AF010	61	Male	5'11”	267	Left	Trauma	Hydraulic
AF011	47	Male	5'9“	224	Left	Accident	Microprocessor
Control participants							
CF023	23	Female	5'8”	140			
CF024	24	Female	6'3“	155			
CF025	24	Female	5'8”	150			
CF026	23	Female	5'6“	130			
CF027	23	Female	5'2”	128			
CF028	25	Male	6'1“	230			
CF029	27	Female	5'0”	100			
CF030	29	Female	5'3“	105			
CF031	21	Male	5'9”	260			
CF032	23	Male	5'10“	145			

We compared the performance of a fall-detection model trained on data from control individuals with that of a model trained on data from the TFA population. We examined how carrying the mobile phone in different locations affected the accuracy of fall detection. To evaluate false positives during everyday life, we assessed the performance of the model when three individuals with TFA carried the phone for a minimum of 48 hours during their daily routine.

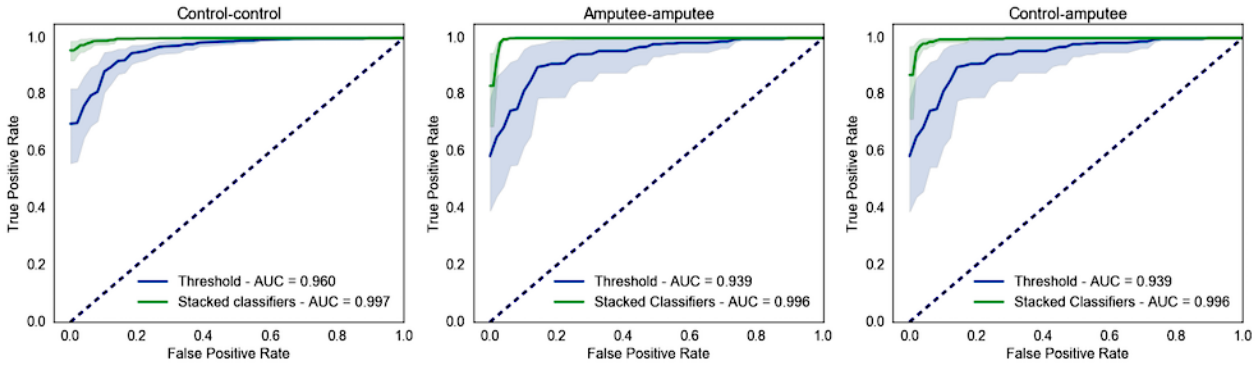
Effect of Population

Our results indicated that a fall-detection classifier trained using data from control participants was able to reliably separate falls from daily activities in individuals with TFA (Figure 4). The performance of this model (control to amputee: AUC mean 0.996, 1.96*standard error of the mean [SEM] 0.004) was not significantly lower than that of a model trained on TFA individuals (amputee to amputee: AUC mean 0.995, SEM 0.004) (Wilcoxon rank-sum test, z=.40, *P*=.69). Fall-detection accuracy for a model trained and tested on non-amputee individuals yielded similar performance (control to control: AUC mean 0.997, SEM 0.003) (z=.65, *P*=.52). Stacking classifiers outperformed the threshold-based classification models in both populations (control to control: z=3.63, *P*<.001; amputee to amputee: z=1.92, *P*=.06; control to amputee: z=2.08, *P*=.04). Table 3 summarizes the fall detection results obtained with each model. Therefore, detecting falls in individuals with TFA could be achieved by only using training falls and activity data from non-amputee individuals.

Table 3. Summary results for models trained and tested on each population (control or amputee). Sensitivity and specificity values represent the optimal point of the ROC curve.

Method and performance metric	Model, mean (1.96 SEM)		
	Control-control	Control-amputee	Amputee-amputee
Stacking			
AUC	0.997 (0.003)	0.996 (0.004)	0.995 (0.004)
Sensitivity	0.979 (0.022)	0.989 (0.017)	0.984 (0.016)
Specificity	0.991 (0.012)	0.968 (0.025)	0.965 (0.022)
Threshold-based			
AUC	0.960 (0.020)	0.939 (0.059)	0.939 (0.059)
Sensitivity	0.915 (0.040)	0.878 (0.097)	0.878 (0.097)
Specificity	0.927 (0.045)	0.922 (0.045)	0.922 (0.045)

Figure 4. Effect of population on model accuracy. Receiver-operator characteristic curves of fall-detection models based on threshold (blue) or using stacked classifiers (green) trained and tested on data from non-amputee individuals (control-control) and individuals with TFA (amputee-amputee), and trained on non-amputee individuals and tested on TFA data (control-amputee). Shaded areas are 95% confidence intervals from bootstrapping.



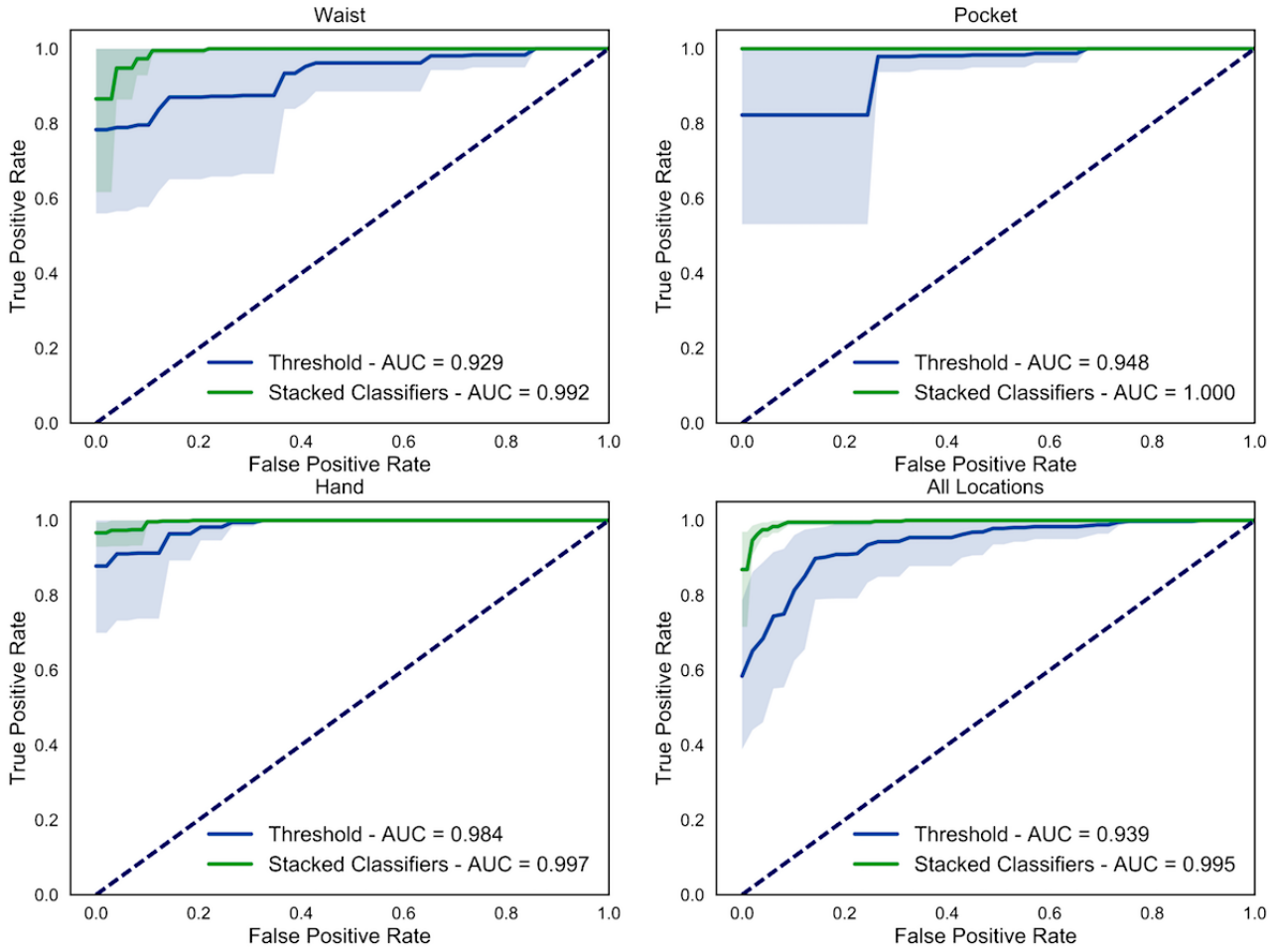
Effect of Location

To assess how the location where the phone was carried affected fall-detection accuracy, we examined the performance of the control-amputee model for each location separately (Figure 5). Carrying the phone in the pocket yielded the highest AUC (mean 1.000, SEM 0.000) followed by hand (mean 0.997, SEM 0.003) and waist (mean 0.992, SEM 0.012); however, no statistically significant differences were found between these values (z=0.37-1.47, *P*=.14-.72). Our stacked classifier model showed less intersubject variability than threshold approaches. A summary of results is reported in Table 4. Regardless of the location, the AUC values suggest good algorithm performance for the stacked classifiers.

Table 4. Summary results for the control-to-amputee model tested on in-laboratory data organized by phone location.

Method and performance metric	Mobile phone location, mean (1.96 SEM)			
	Waist	Pocket	Hand	All
Stacked classifiers				
AUC	0.992 (0.012)	1.000 (0.000)	0.997 (0.003)	0.996 (0.004)
Sensitivity	0.990 (0.017)	1.000 (0.000)	0.989 (0.013)	0.989 (0.017)
Specificity	0.982 (0.033)	1.000 (0.000)	0.980 (0.036)	0.968 (0.025)
Threshold				
AUC	0.929 (0.097)	0.948 (0.065)	0.984 (0.027)	0.939 (0.059)
Sensitivity	0.932 (0.087)	0.979 (0.036)	0.995 (0.010)	0.878 (0.097)
Specificity	0.915 (0.112)	0.934(0.115)	0.939 (0.092)	0.922 (0.045)

Figure 5. Effect of phone location on fall-classification accuracy. Receiver-operator characteristic curves for the control-amputee model organized by test location (green: stacked classifiers; blue: threshold model). Shaded areas are 95% confidence intervals.



Home Data Analysis

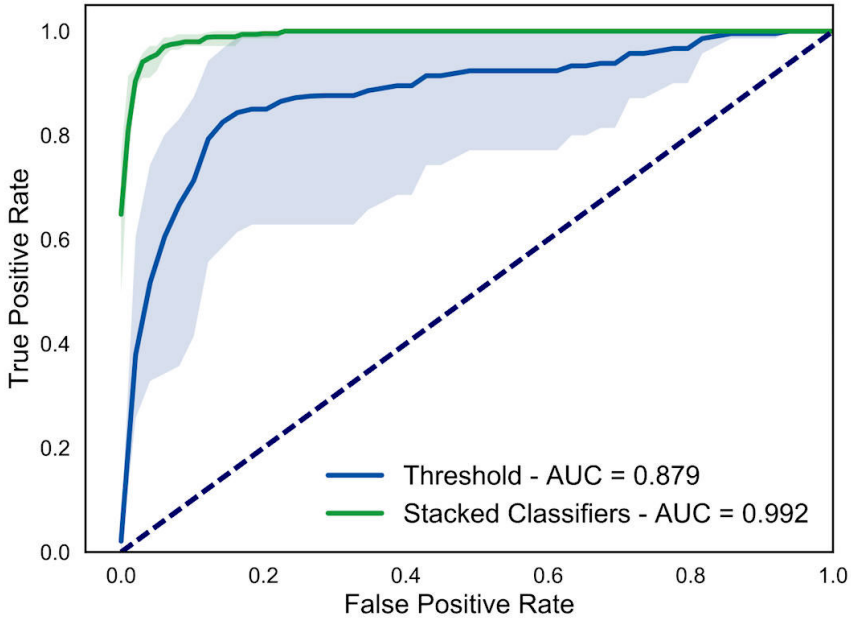
A fall-detection model can be effectively deployed outside of a laboratory setting if the number of false alarms generated is small. Therefore, we tested our model on data from three individuals with TFA who, following the experimental session, carried the mobile phone with them during their daily routine for a period between 2 and 7 days. All participants chose to carry the phone in their pocket and were not given any indication on how to carry the phone otherwise. The participants did not experience a fall during the recording period. Testing the model during the home period allowed us to meaningfully assess the false positive rate.

Performance of our model on the home data was comparable to performance obtained on laboratory data (control-to-amputee model, Figure 6 and Table 5; z=1.03, *P*=.30), with a mean AUC across the three participants of 0.992 (SEM 0.001); sensitivity: mean 0.970, SEM 0.021; specificity: mean 0.950, SEM 0.016. The stacked classifiers model also performed better than the threshold method (z=1.96, *P*=.05), with all results reported in Table 5. We also calculated the false alarms rates for these three participants: at a sensitivity of 90%, there were 1.0, 4.6, and 1.1 false alarms per day (mean 2.2, SD 1.7), whereas for the threshold approach, the false alarm rates were 4.9, 357.0, and 4.4 per day (mean 122.1, SD 166.1), respectively. The high false positive rate for the second participant appears to be the result of both low acceleration magnitude during simulated falls and high acceleration magnitude during at-home activities. Thus, our model can effectively detect falls while keeping its false alarm rate to a reasonable low value when deployed outside of a laboratory-controlled scenario.

Table 5. Summary results for each model on home data.

Performance metric	Method, mean (1.96 SEM)	
	Stacked classifiers	Threshold
AUC	0.992 (0.001)	0.879 (0.121)
Sensitivity	0.970 (0.021)	0.842 (0.188)
Specificity	0.950 (0.016)	0.844 (0.011)

Figure 6. Fall-detection performance on home data. Receiver-operator characteristic curve averaged across the three amputee participants. Data include both the participants' daily routine data and the in-laboratory falls. Shaded areas are 95% confidence intervals.



Discussion

Principal Results

We developed a fall-detection classifier using data collected from the inertial sensors of a mobile phone, carried by non-amputee individuals and by individuals with a TFA. In order to mimic a naturalistic setting, phones were carried in three common locations (pouch, pocket, or hand) and without standardizing the orientation of the phone. We observed no significant effect of the population used for training the model on the fall detection accuracy; a model trained on non-amputee data was as accurate as one trained on our pool of individuals with TFA. Therefore, fall detection can be reliably performed in amputee participants using data from non-amputee participants.

Previous studies have generally used validation on non-amputee participants despite the fact that the clinical population is the real target. However, participants with mobility impairments display different movement patterns from unimpaired individuals, which can affect the accuracy of activity recognition classifiers [28-30]. We pursued the possibility that amputee movements during activities and simulated fall events may have been unique enough to suggest population-specific model training. Our results did not show such a dependency, thus suggesting that a fall-detection model trained on non-amputee individuals can generalize to other clinical populations prone to falls.

We collected falls and activity data as the phone was carried across common locations to mimic a naturalistic scenario. Nevertheless, we observed that the peak acceleration from real-world data exceeded the range of accelerations of the in-laboratory activities. This is not surprising because prior studies also found that in-laboratory activities look different from real-world unstructured behaviors [30]. Thus, falls and activity data during natural use of the phone must be collected to build a fall-detection system that can be deployed in an everyday scenario.

We compared a machine-learning model based on 40 features to a threshold-based approach, which used a single feature (maximum acceleration) to detect whether the motion of the phone constituted a fall event. Our model yielded an average of two false alarms per day versus approximately 122 produced by the threshold model. This result is at least as good as current state-of-the-art fall-detection systems based on a waist-mounted wearable accelerometer [6]. Therefore, combining multiple features through machine learning confers a significant advantage to build a robust phone-based fall detection system.

The purpose of this work was to develop a system that can capture real-world falls with high probability, while reducing the number of false alarms. In addition to the inertial measurement unit (accelerometer, gyroscope) data used for the study, mobile phones can also collect location and weather data, as well as responses to survey questions sent to the participants. This data could be uploaded to a remote server for further analysis, including classification of fall types and the activity preceding the fall. This host of information can be used to build better fall monitoring systems, as well as understanding the context where real falls can occur in amputees.

Limitations

One limitation of this study is the fact that we did not incorporate real-world falls, but rather falls in a controlled setting either initiated by the participant (for slips and trips) or induced by pushing (for lateral falls). Our goal was to collect falls data that approximate real falls, while being practical and safe for our participants. Real falls may have a different movement pattern than simulated falls [31], and algorithms developed on simulated falls can fail when tested on real falls [6]. We will acquire real-world falls data for future work by letting participants carry the mobile phones and report the natural falls when they occur using the Purple Robot app; this

data could then be used to refine the detection algorithm and improve the system reliability over time.

Although two false alarms per day is a reasonably low value, this result might still produce a large number of false positives relative to the total number of falls that can be expected to capture. Additional analysis is necessary to differentiate false positives from actual falls. This could be done by analyzing other sources of information surrounding the fall, such as GPS and activity data. For example, this would distinguish someone who fell and remained on the ground from someone who lost their balance but continued to walk afterwards, or a phone dropped on the ground and then picked up. Such an analysis could be performed either on the phone or on a remote server, depending on its complexity.

A larger feature set or larger clip lengths could also be used to further reduce the number of false positives, and remains to be explored. For example, previous models have used features describing posture and frequency domain features [6,15]. However, the computational complexity has to be balanced against the need for real-time continuous monitoring. Alternatively, a more complex analysis could be also run on a remote server, with the phone filtering for probable falls.

Conclusions

We developed a machine-learning classifier to detect falls in people with lower limb amputations using data from mobile phone inertial sensors. Our results demonstrate that a classifier trained on falls from non-amputee participants can reliably generalize to other populations while the mobile phone is naturally carried in multiple locations. Our approach yields a significant advantage over a threshold-based classifier because it drastically reduces the number of false alarms, which is arguably necessary for fall-detection system to be of practical use. By applying the techniques used here, along with improvements in battery management, we believe that fall detection comes one step closer to improving the interventions performed after individuals with disabilities experience a fall. Currently, we are in the process of collecting real-world falls in individuals with lower limb amputations to test our fall-detection system in everyday life.

Acknowledgments

The authors are thankful to Mark Begale and Christopher J Karr from CBITs for their technical assistance with the Purple Robot phone app. The authors would also like to thank Dr Saninder Kaur, Kelsey Greenoe, Ashley Adamczyk, and Ryan Griesenauer for their assistance during participant recruitment and data collection. This study was funded by the National Institute of Health – NIBIB grant 5 R01 EB019406-04 and by the Max Näder Rehabilitation Technologies and Outcomes Research Center of the Shirley Ryan Ability Lab (formerly Rehabilitation Institute of Chicago).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Activity protocol. The set of activities performed by the subjects to mimic daily functional activities.

[PDF File (Adobe PDF File), 12KB - mhealth_v5i10e151_app1.pdf]

References

1. Fulks J, Fallon F, King W, Shields G, Beaumont N, Ward-Lonergan J. Accidents and falls in later life. Generations Rev 2002;12(3):3.
2. Sixsmith A, Johnson N. A smart sensor to detect the falls of the elderly. IEEE Pervasive Comput 2004 Apr;3(2):42-47. [doi: 10.1109/MPRV.2004.1316817]
3. Kulkarni J, Wright S, Toole C, Morris J, Hirons R. Falls in patients with lower limb amputations: prevalence and contributing factors. Physiotherapy 1996 Feb;82(2):130-136. [doi: 10.1016/S0031-9406(05)66968-4]
4. Miller WC, Speechley M, Deathe B. The prevalence and risk factors of falling and fear of falling among lower extremity amputees. Arch Phys Med Rehabil 2001 Aug;82(8):1031-1037. [doi: 10.1053/apmr.2001.24295] [Medline: 11494181]
5. Gurley RJ, Lum N, Sande M, Lo B, Katz MH. Persons found in their homes helpless or dead. N Engl J Med 1996 Jun 27;334(26):1710-1716. [doi: 10.1056/NEJM199606273342606] [Medline: 8637517]
6. Bagalà F, Becker C, Cappello A, Chiari L, Aminian K, Hausdorff JM, et al. Evaluation of accelerometer-based fall detection algorithms on real-world falls. PLoS One 2012;7(5):e37062 [FREE Full text] [doi: 10.1371/journal.pone.0037062] [Medline: 22615890]
7. Dai J, Bai X, Yang Z, Shen Z, Xuan D. PerFallD: a pervasive fall detection system using mobile phones. In: Pervasive Computing and Communications Workshops (PERCOM Workshops). 2010 Presented at: 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops); Mar 29-Apr 2, 2010; Mannheim, Germany p. 292-297. [doi: 10.1109/PERCOMW.2010.5470652]
8. Lee RY, Carlisle AJ. Detection of falls using accelerometers and mobile phone technology. Age Ageing 2011 Nov;40(6):690-696. [doi: 10.1093/ageing/afr050] [Medline: 21596711]

9. Fahmi P, Viet V, Deok-Jai C. Semi-supervised fall detection algorithm using fall indicators in smartphone. In: Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication. 2012 Presented at: ICUIMC '12 6th International Conference on Ubiquitous Information Management and Communication; Feb 20-22, 2012; Kuala Lumpur p. 122. [doi: 10.1145/2184751.2184890]
10. Yavuz G, Kocak M, Ergun G, Alemdar H, Yalcin H, Incel O, et al. A smartphone based fall detector with online location support. In: Proceedings of PhoneSense. 2010 Presented at: PhoneSense'10 First International Workshop on Sensing for App Phones; Nov 2, 2010; Zurich p. 31-35 URL: http://sensorlab.cs.dartmouth.edu/phonesense/papers/Yavuz-fall.pdf
11. Mellone S, Tacconi C, Schwickert L, Klenk J, Becker C, Chiari L. Smartphone-based solutions for fall detection and prevention: the FARSEEING approach. Z Gerontol Geriatr 2012 Dec;45(8):722-727. [doi: 10.1007/s00391-012-0404-5] [Medline: 23184298]
12. Habib MA, Mohktar MS, Kamaruzzaman SB, Lim KS, Pin TM, Ibrahim F. Smartphone-based solutions for fall detection and prevention: challenges and open issues. Sensors (Basel) 2014 Apr 22;14(4):7181-7208 [FREE Full text] [doi: 10.3390/s140407181] [Medline: 24759116]
13. Delahoz YS, Labrador MA. Survey on fall detection and fall prevention using wearable and external sensors. Sensors (Basel) 2014 Oct 22;14(10):19806-19842 [FREE Full text] [doi: 10.3390/s141019806] [Medline: 25340452]
14. Sposaro F, Tyson G. iFall: an Android application for fall monitoring and response. In: Proceedings of the Engineering in Medicine and Biology Society. 2009 Presented at: EMBC 2009 Annual International Conference of the IEEE; Sep 3-6, 2009; Minneapolis, MN p. 6119-6122. [doi: 10.1109/IEMBS.2009.5334912]
15. Albert MV, Kording K, Herrmann M, Jayaraman A. Fall classification by machine learning using mobile phones. PLoS One 2012 May;7(5):e36556 [FREE Full text] [doi: 10.1371/journal.pone.0036556] [Medline: 22586477]
16. Tolkiehn M, Atallah L, Lo B, Yang G. Direction sensitive fall detection using a triaxial accelerometer and a barometric pressure sensor. Conf Proc IEEE Eng Med Biol Soc 2011;2011:369-372. [doi: 10.1109/IEMBS.2011.6090120] [Medline: 22254325]
17. Luštrek M, Kaluža B. Fall detection and activity recognition with machine learning. Informatica 2009;33(2):197-204.
18. Aziz O, Klenk J, Schwickert L, Chiari L, Becker C, Park EJ, et al. Validation of accuracy of SVM-based fall detection system using real-world fall and non-fall datasets. PLoS One 2017 Jul;12(7):e0180318 [FREE Full text] [doi: 10.1371/journal.pone.0180318] [Medline: 28678808]
19. Medrano C, Igual R, Plaza I, Castro M. Detecting falls as novelties in acceleration patterns acquired with smartphones. PLoS ONE 2014 Apr 15;9(4):e94811. [doi: 10.1371/journal.pone.0094811]
20. Tinetti M, Doucette J, Claus E, Marottoli R. Risk factors for serious injury during falls by older persons in the community. J Am Geriatr Soc 1995 Nov;43(11):1214-1221. [doi: 10.1111/j.1532-5415.1995.tb07396.x]
21. Schueller SM, Begale M, Penedo FJ, Mohr DC. Purple: a modular system for developing and deploying behavioral intervention technologies. J Med Internet Res 2014 Jul 30;16(7):e181 [FREE Full text] [doi: 10.2196/jmir.3376] [Medline: 25079298]
22. Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one? Mach Learn 2004 Mar;54(3):255-273. [doi: 10.1023/B:MACH.0000015881.36452.6e]
23. Chang C, Lin C. LIBSVM. ACM Trans Intell Syst Technol 2011 Apr 01;2(3):1-27. [doi: 10.1145/1961189.1961199]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 13-17, 2016; San Francisco p. 785-794. [doi: 10.1145/2939672.2939785]
25. Krishnapuram B, Carin L, Figueiredo MAT, Hartemink AJ. Sparse multinomial logistic regression: fast algorithms and generalization bounds. IEEE Trans Pattern Anal Mach Intell 2005 Jun;27(6):957-968. [doi: 10.1109/TPAMI.2005.127] [Medline: 15943426]
26. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Statistical Soc B 2005 Apr;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
27. Bourke AK, O'Brien JV, Lyons GM. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. Gait Posture 2007 Jul;26(2):194-199. [doi: 10.1016/j.gaitpost.2006.09.012] [Medline: 17101272]
28. Albert MV, Toledo S, Shapiro M, Kording K. Using mobile phones for activity recognition in Parkinson's patients. Front Neurol 2012;3:158 [FREE Full text] [doi: 10.3389/fneur.2012.00158] [Medline: 23162528]
29. Lonini L, Gupta A, Kording K, Jayaraman A. Activity recognition in patients with lower limb impairments: do we need training data from each patient? Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:3265-3268. [doi: 10.1109/EMBC.2016.7591425] [Medline: 28269004]
30. O'Brien MK, Shawen N, Mummidisetty CK, Kaur S, Bo X, Poellabauer C, et al. Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting. J Med Internet Res 2017 May 25;19(5):e184 [FREE Full text] [doi: 10.2196/jmir.7385] [Medline: 28546137]
31. Klenk J, Becker C, Lieken F, Nicolai S, Maetzler W, Alt W, et al. Comparison of acceleration signals of simulated and real-world backward falls. Med Eng Phys 2011 Apr;33(3):368-373. [doi: 10.1016/j.medengphy.2010.11.003] [Medline: 21123104]

Abbreviations

AUC: area under the curve
GPS: Global Positioning System
LOSOCV: leave-one-subject-out cross-validation
ROC: receiver-operator characteristic
SEM: standard error of the mean
TFA: transfemoral amputation

Edited by G Eysenbach; submitted 13.06.17; peer-reviewed by R Robinson, C Becker, MH Cameron; comments to author 01.08.17; revised version received 08.08.17; accepted 10.08.17; published 11.10.17.

Please cite as:
Shawen N, Lonini L, Mummidisetty CK, Shparii I, Albert MV, Kording K, Jayaraman A
Fall Detection in Individuals With Lower Limb Amputations Using Mobile Phones: Machine Learning Enhances Robustness for Real-World Applications
JMIR Mhealth Uhealth 2017;5(10):e151
URL: http://mhealth.jmir.org/2017/10/e151/
doi:10.2196/mhealth.8201
PMID:29021127

Original Paper

iPhone Sensors in Tracking Outcome Variables of the 30-Second Chair Stand Test and Stair Climb Test to Evaluate Disability: Cross-Sectional Pilot Study

Gautam Adusumilli[1*], BA; Solomon Eben Joseph[1*]; Michael A Samaan[1], PhD; Brooke Schultz[2], MS; Tijana Popovic[1], MS; Richard B Souza[1,3], PT, PhD; Sharmila Majumdar[1], PhD

[1]Musculoskeletal Quantitative Imaging Research Group, Department of Radiology and Imaging, University of California San Francisco, San Francisco, CA, United States

[2]Human Performance Center, Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, CA, United States

[3]Human Performance Center, Department of Physical Therapy and Rehabilitation Science, University of California San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Gautam Adusumilli, BA
Musculoskeletal Quantitative Imaging Research Group
Department of Radiology and Imaging
University of California San Francisco
Lobby 6, Suite 350
185 Berry St
San Francisco, CA, 94107
United States
Phone: 1 919 576 3243
Email: gautam.adusumilli@wustl.edu

Abstract

Background: Performance tests are important to characterize patient disabilities and functional changes. The Osteoarthritis Research Society International and others recommend the 30-second Chair Stand Test and Stair Climb Test, among others, as core tests that capture two distinct types of disability during activities of daily living. However, these two tests are limited by current protocols of testing in clinics. There is a need for an alternative that allows remote testing of functional capabilities during these tests in the osteoarthritis patient population.

Objective: Objectives are to (1) develop an app for testing the functionality of an iPhone's accelerometer and gravity sensor and (2) conduct a pilot study objectively evaluating the criterion validity and test-retest reliability of outcome variables obtained from these sensors during the 30-second Chair Stand Test and Stair Climb Test.

Methods: An iOS app was developed with data collection capabilities from the built-in iPhone accelerometer and gravity sensor tools and linked to Google Firebase. A total of 24 subjects performed the 30-second Chair Stand Test with an iPhone accelerometer collecting data and an external rater manually counting sit-to-stand repetitions. A total of 21 subjects performed the Stair Climb Test with an iPhone gravity sensor turned on and an external rater timing the duration of the test on a stopwatch. App data from Firebase were converted into graphical data and exported into MATLAB for data filtering. Multiple iterations of a data processing algorithm were used to increase robustness and accuracy. MATLAB-generated outcome variables were compared to the manually determined outcome variables of each test. Pearson's correlation coefficients (PCCs), Bland-Altman plots, intraclass correlation coefficients (ICCs), standard errors of measurement, and repeatability coefficients were generated to evaluate criterion validity, agreement, and test-retest reliability of iPhone sensor data against gold-standard manual measurements.

Results: App accelerometer data during the 30-second Chair Stand Test (PCC=.890) and gravity sensor data during the Stair Climb Test (PCC=.865) were highly correlated to gold-standard manual measurements. Greater than 95% of values on Bland-Altman plots comparing the manual data to the app data fell within the 95% limits of agreement. Strong intraclass correlation was found for trials of the 30-second Chair Stand Test (ICC=.968) and Stair Climb Test (ICC=.902). Standard errors of measurement for both tests were found to be within acceptable thresholds for MATLAB. Repeatability coefficients for the 30-second Chair Stand Test and Stair Climb Test were 0.629 and 1.20, respectively.

Conclusions: App-based performance testing of the 30-second Chair Stand Test and Stair Climb Test is valid and reliable, suggesting its applicability to future, larger-scale studies in the osteoarthritis patient population.

(JMIR Mhealth Uhealth 2017;5(10):e166) doi:10.2196/mhealth.8656

KEYWORDS

osteoarthritis; telemedicine; mobile phone; mobile apps; algorithms; medical informatics

Introduction

Osteoarthritis is the most prevalent chronic condition of the joints, affecting approximately 30 million Americans and one in every two adults during their lifetimes [1]. Osteoarthritis most commonly occurs in the weight-bearing hip and knee joints and is the consequence of a progressive breakdown process of articular cartilage, joint capsule, and ligaments; synovial tissue inflammation; and subchondral bone sclerosis [2-5]. Patients diagnosed with osteoarthritis experience a decline in physical function and often report difficulty performing activities of daily living (ADL) [6]. Recent work has shown that osteoarthritis patients place increased compensatory stress on their unaffected weight-bearing joints during ADL, a vicious cycle that makes them susceptible to further damage and worsened severity of disease [7-9].

It is imperative for the development of a treatment plan to identify osteoarthritis-induced deteriorations in physical function and each patient's respective compensatory mechanisms. In 2013, a battery of five tests simulating ADL was recommended by the Osteoarthritis Research Society International (OARSI) for this purpose [10]. The 30-second Chair Stand Test (CST), 40-meter Fast-Paced Walk Test, and Stair Climb Test (SCT) were listed as the three minimal core set of performance-based tests; the 6-Minute Walk Test (6MWT) and Timed Up and Go were listed as noncore tests. Outcome variables from the CST have since been found to be significant predictors of pain, self-reported disability, fall risk, and decreased lower-extremity strength in osteoarthritis patients [11,12]. Similarly, strong associations have been found between the SCT, self-reported walking limitation, and quadriceps and hamstring strength [13]. The simplicity of both the CST and SCT makes possible testing in both clinical and residential settings.

There is precedent in the use of iPhones to quantify outcome variables during the 6MWT in congenital heart failure patients [14]. The ubiquity of iPhones and their built-in sensors may allow for the widespread extrapolation of this utility to the CST and SCT. The primary outcome variables of (1) CST: total number of sit-to-stand cycles and (2) SCT: total ascent and descent duration must be captured by iPhones for this extrapolation to be successful [11,12,15].

Our long-term proposal is to develop an iOS app accessible to the world population, with the ability to instruct osteoarthritis patients in any setting through the CST and SCT, capture clinical outcome data, and export results to clinicians and researchers. As a precursor to this, we developed a prototype app to demonstrate proof of concept and evaluated two iOS sensors of interest: the linear accelerometer and the gravity sensor. We discuss in this paper our experience working with these sensors and report on the criterion validity and test-retest reliability of CST and SCT outcome variable data captured by the app as compared to gold-standard measurements during each test.

Methods

iPhone Sensor Triaxial Alignment and Phone Placement

The iPhone accelerometer has a preprogrammed triaxial alignment, as depicted in Figure 1. Standing up out of a chair during CST is considered Z acceleration in the 3D-world coordinate system, but the phone may experience X, Y, Z, or a combination of accelerations depending on its orientation.

The CST protocol requires subjects to cross their arms across their chest as they stand up out of a chair, so we opted to conduct testing at an analogous phone location and orientation: holding the phone vertically in hand against the chest with arms crossed (see Figure 2).This placement gives Y acceleration peak output during each stand and would make widespread CST testing a practicality, without a requirement for extra equipment.

The gravity sensor was used for the SCT because of its ability to sense changes in gravitational moments along the three axes. We decided to place the phone in the subject's pants pocket during the SCT. When the subject is standing upright, the phone is vertical in the pocket and feels gravity on its Y-axis; but as the leg is lifted to take the first step, the phone becomes horizontal and the phone feels gravity on its Z-axis, with a simultaneous decrease to zero on the Y-axis. We sought to isolate these features during the SCT to determine the primary outcome variable, total duration of the test. Validation of data obtained at this pocket location could make widespread testing of the SCT a practicality as well.

Figure 1. A 3D depiction of an iPhone's triaxial coordinate system. Notice that both axis and direction are contingent on the phone's orientation during movement.

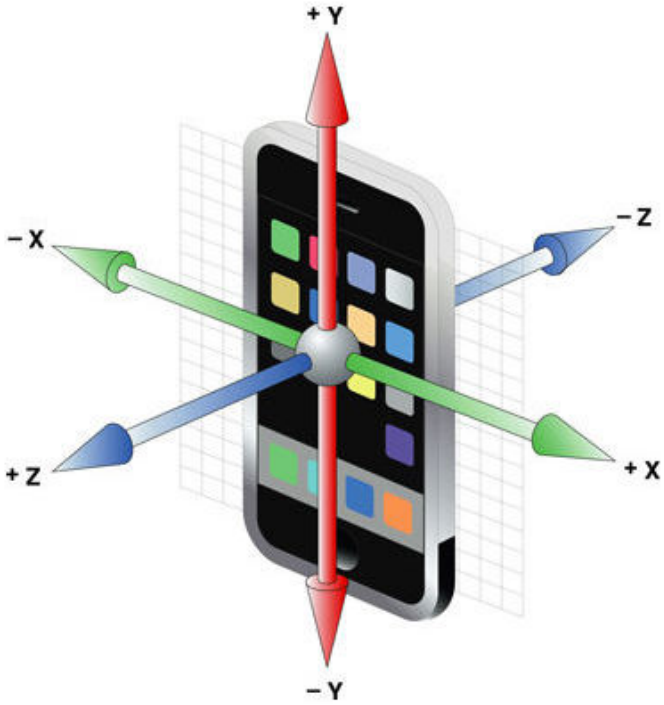


Figure 2. A representation of the phone's orientation and location during Chair Stand Test testing. An iPhone with its screen facing inward and home button downward would experience +Y acceleration to represent upward vertical movement and –Z acceleration to represent the subject moving forward during the postural transition of standing.

Participant Recruitment

Patients were recruited from the University of California San Francisco (UCSF) Center of Research Translation for the Study of Osteoarthritis. Healthy controls were recruited internally from the Musculoskeletal Quantitative Imaging Research Group at UCSF. Sample size was targeted to be a number between 24 and 50 based on recommendations for pilot studies in clinical research [16,17]. Exclusion criteria for controls were confounders that, in the opinion of the tester, could affect the postural transition of standing up (eg, lower extremity orthopedic conditions, morbid obesity, and pregnancy). Patients and controls consented after the study was ethically approved by the UCSF Human Research Protection Office and Institutional Review Board.

Prototype App Development and Functionality

The prototype app was designed with Swift 3.0 on the XCode 8.3.3 platform (Apple Inc). After using an existing app on the iOS app store that graphically displayed accelerometer and gravity sensor data to determine the optimal phone placement and sensor data collection, we integrated relevant sensor data collection into a user-friendly app with account capabilities and testing prompts to allow for longitudinal participation.

We used the Core Motion framework (Apple Inc) to give the app the ability to collect raw accelerometer data reflecting phone motion. We used FirebaseDatabase and FirebaseAuth frameworks (Google Inc) to set up data storage in Google Firebase, a cloud-hosted real-time database that can facilitate user authentication for iOS apps.

Subject identifier fields requested in the app included the following: *Name*, *Date of Birth*, *Email*, *Password*, and *Subject Classification* (Osteoarthritis, Femoroacetabular Impingement, and Healthy Control). In addition, subjects were prompted prior to the Stair Climb Test to choose the number of steps they will ascend and descend from a range of 8-15 steps.

The 30-second data recording period of the CST was started using a toggle button, with a 3-second latency to allow the subject to cross their arms and orient the phone. This latency is coupled with a countdown on the screen and beeps of various tones to signal the inception of the 30-second countdown.

The recording period of the SCT was also initiated with a toggle button and a 3-second latency to allow the subject to put the phone in their pants pocket, coupled with the same visual and audial cues. The recording period was set to be 45 seconds, with an option to take the phone out of the pocket and end data collection if the test was completed early. Sampling rate was set constant at 100 Hz for both tests, but small fluctuations were observed in the sampling frequency between 100 Hz and 101 Hz as the app collected data.

Performance Testing Protocol

Testing was conducted in various sites at the UCSF Departments of Orthopaedic Surgery and Radiology.

Chair Stand Test

Prior to the CST, a chair of 18 inches in height was placed against a wall to ensure stability. The tester provided an iPhone 6 with the prototype app and requested subjects to hold the phone vertically in their right hand during testing, with the home button facing down and screen facing inward. Subjects were asked to repeat cycles of standing up and sitting down completely at a self-selected speed for 30 seconds during the test.

Each subject pressed the start button and positioned the phone quickly across the chest during the 3-second latency period. At the sound of the start buzzer, subjects began standing and sitting cycles, with the tester manually counting the total number of sit-to-stand cycles. An end buzzer from the app signaled the end of data collection. The entire protocol was repeated for two trials in all subjects. An optional resting period was given between trials as required to reduce fatigue as a confounder. Upon completion of both trials, data were automatically pushed to Google Firebase in real time as JSON files.

Stair Climb Test

A stairway of 12 steps was located at each of the UCSF testing locations. The tester provided an iPhone with the prototype app and instructed each subject to choose the leg they planned to take their first step with. Subjects were requested to climb the steps and descend at a self-selected speed.

Each subject pressed the start button and placed the phone in the pocket of the leg they chose to take their first step with. At the sound of the start buzzer, the tester started a stopwatch and the subject proceeded to ascend the stairs. At the end of the descent, after both feet had touched the base of the steps, the stopwatch was stopped. The subject was instructed to take the phone out of their pocket and manually end data collection if the test took less than 45 seconds. The entire protocol was repeated for two trials in all subjects, with a 120-second resting period preset in the app between trials as required. Data was automatically pushed to Google Firebase after completion of the test.

Data Processing

Overview

After exporting the raw data from each test out of Google Firebase in a JSON file format, we used an online conversion tool to change the file format from JSON to CSV. The data from each test were then imported into MATLAB (The MathWorks, Inc) for data processing.

Chair Stand Test

In MATLAB, graphs of acceleration in the Y-axis were phase-shifted by subtracting the mean to compensate for the effects of gravity and time-normalized to exactly 3000 frames (30 seconds at 100 Hz) to remediate variation in data collection frequency. A 3 Hz low-pass, fourth-order Butterworth filter and magnitude thresholds were applied to eliminate extraneous peaks due to recoil noise.

After completion of graph processing, local extrema were automatically identified, using a threshold of 0.4 m/s^2, and labeled in MATLAB as sit-to-stand postural transitions of the individual. The total number of repetitions identified for each trial was juxtaposed with manual counts from testing.

Stair Climb Test

The data were phase-shifted in MATLAB by subtracting the mean and a 3 Hz low-pass, fourth-order Butterworth filter was applied to eliminate extraneous data points due to recoil noise. The derivative of each graph was taken to determine change in acceleration. Test graphical features were determined using a threshold of 0.03 m/s^3 for local extrema. Since a peak in the graph occurred for alternating steps due to the placement of the phone adjacent to one limb, we observed six peaks during the ascent and six during the descent of a 12-step test. The twelfth peak was set as the end marker for the test and the corresponding frame number for this peak was outputted as the outcome variable, test duration, by dividing the frame number by the data collection frequency of 100 Hz.

Statistical Analysis

Validation of Chair Stand Test Outcome Variable: Total Number of Sit-to-Stand Cycles

Observed sit-to-stand count was compared to +Y-acceleration peak count outputted by MATLAB to validate the iPhone's ability to capture the sit-to-stand feature of each cycle. The average of trials was taken for each subject to obtain average MATLAB count and average manual count for that subject. These values were compared across all subjects to obtain average absolute difference between MATLAB and human counts. Pearson's correlation coefficient (PCC) was determined in Microsoft Excel between the variables of average MATLAB count and average manual count.

To determine the intraclass correlation coefficient (ICC), the difference between MATLAB and human count for each individual trial was firstly computed using Microsoft Excel. The average of differences across trials was taken for each individual subject to obtain test-retest means. The average and standard deviation of these test-retest means (SDtest-retest) was computed and SPSS Statistics version 23 (IBM Corp) was used to calculate the ICC.

The standard error of measurement (SEM) was calculated as follows:

$$SEM = SDtest\text{-}retest \times \sqrt{(1\text{-}ICC)} \quad (1)$$

SEM was then converted into percentage by dividing by the average value of the manual count. SEM less than 10% was set as the threshold for the acceptability of MATLAB calculations [18].

Repeatability coefficients (CRs) were calculated as follows:

$$CR = 1.96 \times SEM \times \sqrt{2} \quad (2)$$

A Bland-Altman plot with 95% limits of agreement was then created using SPSS to observe the agreement and distribution of differences between app-derived and manual sit-to-stand count.

Validation of Stair Climb Test Outcome Variable: Total Ascent + Descent Duration

Stopwatch time of the SCT test was compared to app-derived duration from MATLAB to validate the iPhone's ability to detect features of the test that enable isolation of the testing interval within the data. Average absolute difference, PCC, SDtest-retest, ICC, and SEM were determined as they were for the CST analysis.

A Bland-Altman plot with 95% limits of agreement was created to evaluate the distribution of differences between app-derived test duration and stopwatch time.

Results

Chair Stand Test and Stair Climb Test Graphical Waveforms

A total of 3 patients (age >50 years; 1 male, 2 female) and 21 healthy controls (age 17-60 years; 12 male, 9 female) participated in testing. Data collection was completely objective and internal controls had a great range of physical capabilities. CST graphical waveforms were characterized by Y peaks and troughs for each sit-to-stand and stand-to-sit repetition, respectively (see Figure 3). Z troughs were observed for the anterior movement during a sit-to-stand repetition and Z peaks for the posterior movement during stand-to-sit repetition. X-axis graphical features were observed if the phone was not held exactly vertical during a trial.

SCT waveforms contained features exclusive to the stair ascent and descent components of the test (see Figure 4). Each step climbed or descended by the limb adjacent to the phone is seen as a decrease in Y acceleration and an increase in Z acceleration to reflect the change of gravitational moments on the phone. The magnitude of these changes was observed to be greater during the stair ascent than the stair descent.

Figure 3. A graphical representation of the 30-second Chair Stand Test as captured by an iOS-based linear accelerometer. Y-axis features represent the postural transitions of sitting and standing and Z-axis features represent anterior and posterior movement during the postural transition. acc: acceleration.

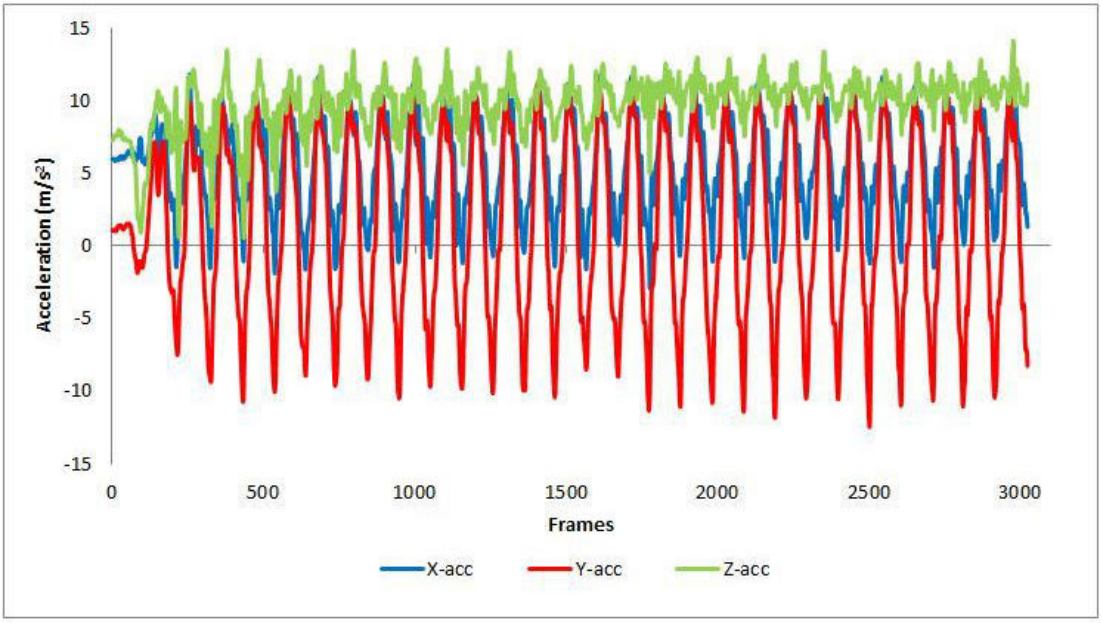
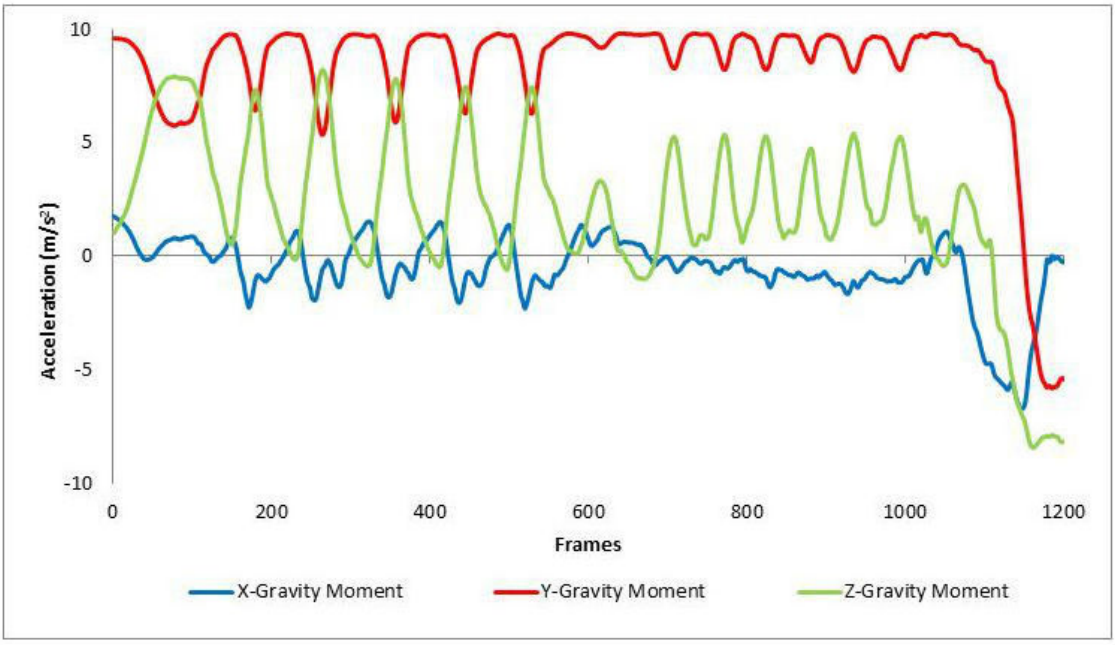


Figure 4. A graphical representation of the Stair Climb Test as captured by an iOS-based gravity sensor. Y-axis features represent changes in gravitational moments during stair ascent and descent.



App-Derived Chair Stand Test Versus Human-Observed Chair Stand Test

A strong Pearson's correlation (see Figure 5) and intraclass correlation was found between the MATLAB count of Y peaks in the CST graphical data and the human count of stands during testing (see Table 1). The SEM was found to be well within the threshold set for acceptability of MATLAB measurements.

This concordance and low variance was observed for all subjects regardless of sit-to-stand count. In the Bland-Altman plot, 96% of points fell within the lines of agreement with an even distribution around the line of mean difference at -0.63 (see Figure 6).

Table 1. Descriptive statistics of Chair Stand Test and Stair Climb Test with comparisons between app-derived data and human-observed data.

Statistic	Chair Stand Test, stands	Stair Climb Test, seconds
Average human count	19.3	10.9
Average MATLAB count	18.6	9.2
Average absolute difference	0.7	1.7
Pearson's correlation coefficient	.890	.865
Average test-retest	1.38	1.77
SDtest-retest[a]	1.27	1.32
Intraclass correlation coefficient	.968	.902
Standard error of measurement	0.227	0.433
Standard error of measurement, %	1.12	3.94
Repeatability coefficient	0.629	1.20

[a]SDtest-retest: standard deviation of test-retest means.

Figure 5. Regression between MATLAB computation and manual count in calculation of 30-second Chair Stand Test repetitions (n=24; Pearson's correlation coefficient=.890).

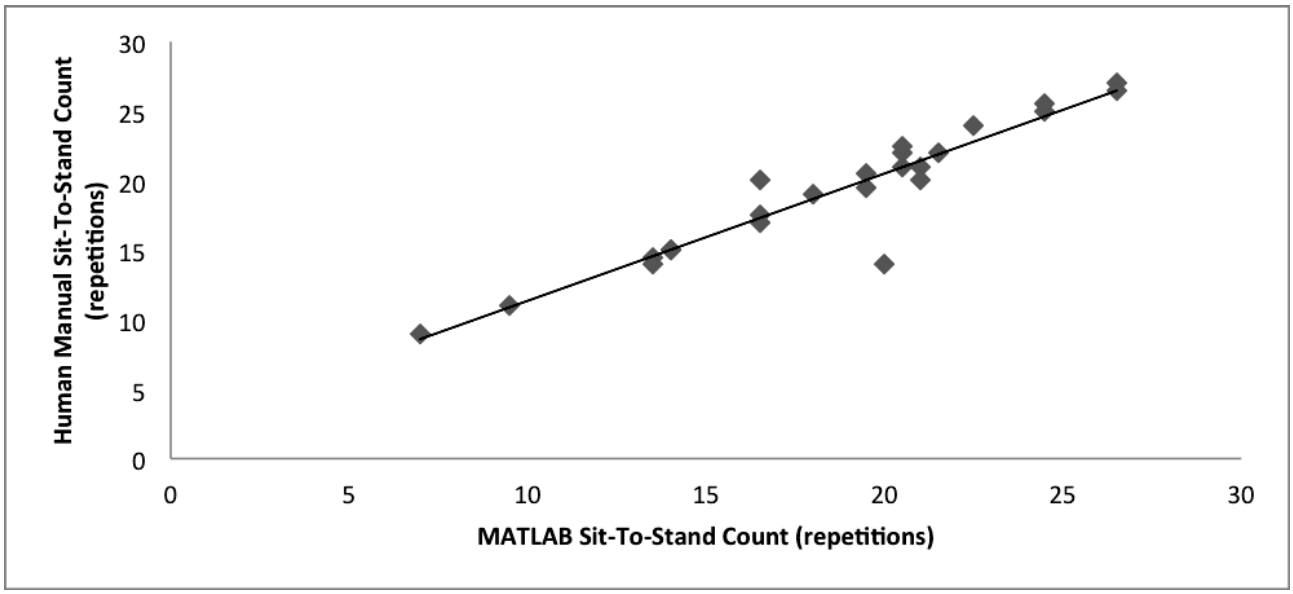
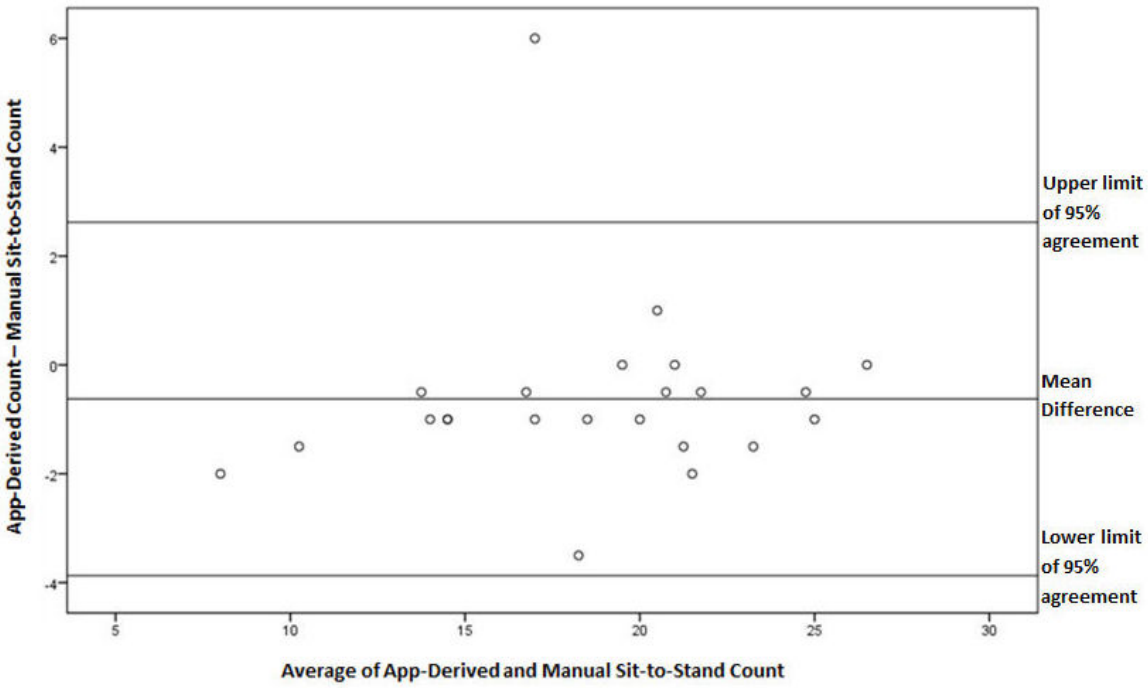


Figure 6. A Bland-Altman plot of differences between app-derived and manual sit-to-stand count during the 30-second Chair Stand Test. Line of mean difference is at -0.63 and upper and lower limits of 95% agreement are at 2.62 and -3.87, respectively.



App-Derived Stair Climb Test Versus Stopwatch-Derived Stair Climb Test

A strong Pearson's correlation and intraclass correlation were observed between MATLAB output test duration and stopwatch-derived duration (see Figure 7 and Table 1). The SEM was within the acceptability threshold and the variance of data appeared to be low between lower and higher test durations (see Figure 7 and Table 1). A total of 95% of data points in the Bland-Altman plot fell within the lines of agreement with a relatively even distribution around the line of mean difference at -1.71 (see Figure 8).

Figure 7. Regression between MATLAB-generated test duration and stopwatch-measured time during the 12-step Stair Climb Test (n=21; Pearson's correlation coefficient=.865).

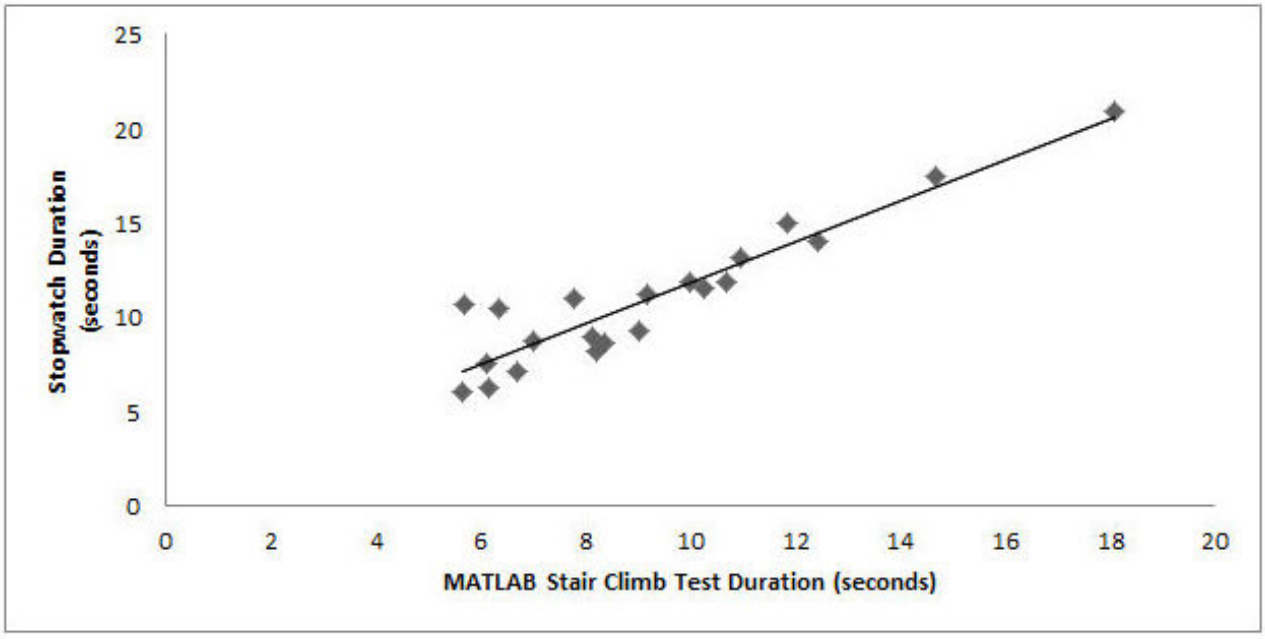
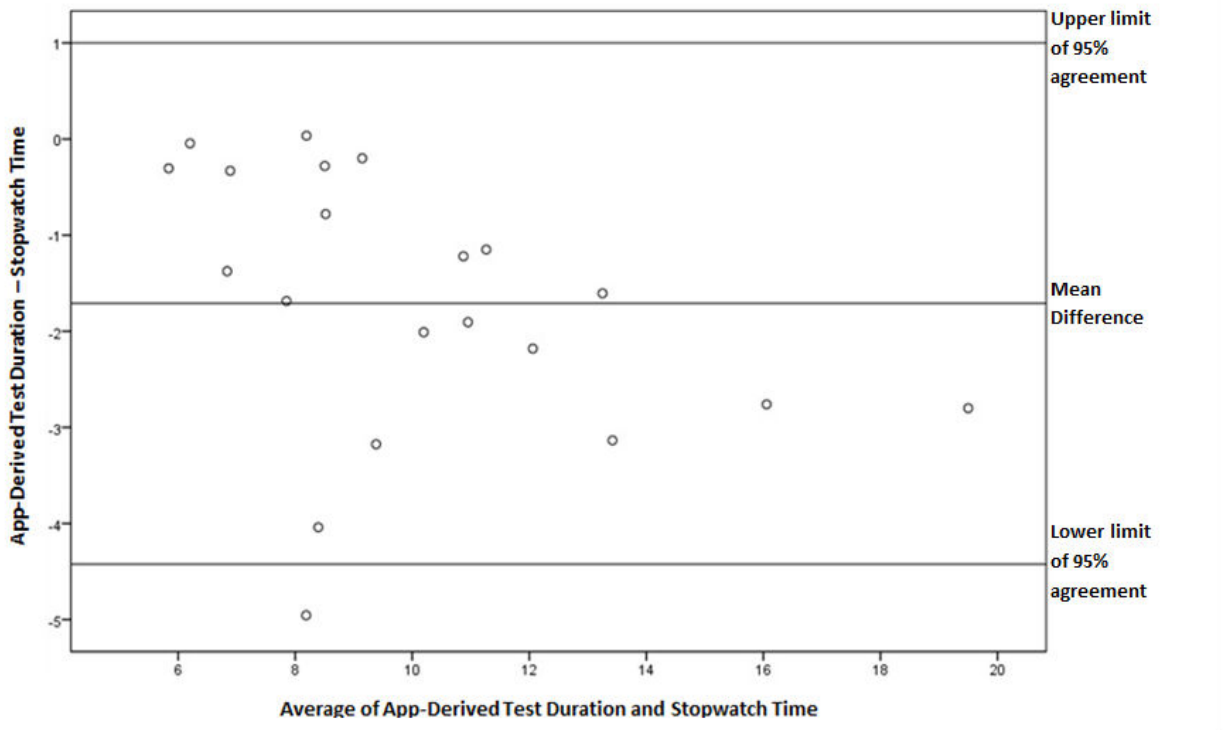


Figure 8. A Bland-Altman plot of differences between app-derived test duration and stopwatch time during the Stair Climb Test. Line of mean difference is at -1.71 and upper and lower limits of 95% agreement are at 1.00 and -4.42, respectively.



Discussion

Principal Findings

Our findings demonstrate that CST and SCT data obtained from iPhone sensors can accurately capture the primary outcome variable from each test. The distinct and consistent data features from each test made it possible to automate data processing using MATLAB. This resulted in very strong values of PCC and ICC and conclusive Bland-Altman plots between manually obtained data and app-derived data, thus establishing criterion validity, test-retest reliability, and agreement. The SEMs for both tests were well below the threshold set in literature, further suggesting the accuracy of our MATLAB algorithm at automating computation of the outcome variables. The CRs further complement the findings of agreement between the app data and the gold-standard manual measurements.

These results have great clinical implication, especially as performance testing using current gold-standard measurements has become a topic of increased interest since OARSI's recommendations were made public in 2013 [10]. The inefficiency of current testing is often complemented by the fact that the data collected in closed clinical settings may not be representative of a patient's disability during ADL in their community. There have been previous efforts at the UCSF Department of Cardiology to make remote testing possible utilizing the ubiquity of iPhones [14]. However, the results are confined to one performance test, the 6MWT, and one iPhone tool, the motion sensor. Though the 6MWT is also a part of the battery of five tests suggested by OARSI, it was not listed as a core test and thus was not chosen as a focus of our study.

Our findings on the two core tests of CST and SCT have greater significance in the osteoarthritis community. Recent papers have reported on their relevancy and importance toward tracking clinical markers of lower extremity strength and fall risk, and patient-reported outcomes such as walking limitation, pain, and disability [11-13]. Furthermore, both tests require rapid hip and knee flexion and extension and a larger range of joint motion than walking tests such as the 6MWT. Because these factors tend to be impaired early in osteoarthritis disease progression, longitudinal examination of these tests may help clinicians predict deteriorations and ultimately develop a better care plan [19,20].

Most subjects in our study reported that they would use the app to test longitudinally at home. While the 6MWT-app study included the requirement of a hip holster for one phone placement location, neither of our phone locations require equipment [14]. This adds to the simplicity of our testing protocol and increases the likelihood of acquiring a large sample of subjects from remote testing locations. On the clinical side, automated data processing with our MATLAB codes could make it practical to longitudinally evaluate and track such a large subject sample. Collectively, this could meet the goals of increasing the frequency of assessments in community settings, decreasing the requirement for clinical visits, and providing a streamlined database of longitudinal physical function data from osteoarthritis patients across the world.

Limitations and Future Directions

Limitations of our study include a low sample size and an overrepresentation of subjects within a younger age group, not well represented in osteoarthritis. Nonetheless, physical abilities varied sufficiently enough to determine iPhone sensor functionality across a large range of outcome variable

magnitudes. The SCT in general appeared to be more difficult to accurately capture than the CST, but given that stopwatch times were consistently overestimated compared to MATLAB-derived times, this may be largely due to human reaction time as a confounder in our study.

The absence of at-home testing within our study can also be considered a limitation. Future studies may be required to definitively establish the validity and reliability of at-home app testing in the presence of external raters and the compliance rate of osteoarthritis patients who receive the app to test longitudinally at home without external raters. Should both studies prove to be successful, a large-scale study focused on establishing minimally clinical importance difference for MATLAB-computed counts of CST and SCT in osteoarthritis patients may be warranted.

Conclusions

An app utilizing the iPhone's accelerometer and gravity sensor might be a good alternate to accurately and consistently obtain outcome variables during the CST and SCT, respectively. The ease of protocol, lack of adverse events during testing, and ability to automate data extraction collectively suggest the preliminary applicability of iPhones as a safe and reliable tool for widespread and longitudinal performance testing in the future.

Acknowledgments

This research was supported by National Institutes of Health grants (R01AR069006 and F32AR069458).

Conflicts of Interest

None declared.

References

1. Cisternas MG, Murphy L, Sacks JJ, Solomon DH, Pasta DJ, Helmick CG. Alternative methods for defining osteoarthritis and the impact on estimating prevalence in a US population-based survey. Arthritis Care Res (Hoboken) 2016 May;68(5):574-580 [FREE Full text] [doi: 10.1002/acr.22721] [Medline: 26315529]
2. Brody LT. Knee osteoarthritis: Clinical connections to articular cartilage structure and function. Phys Ther Sport 2015 Nov;16(4):301-316. [doi: 10.1016/j.ptsp.2014.12.001] [Medline: 25783021]
3. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: An update with relevance for clinical practice. Lancet 2011 Jun 18;377(9783):2115-2126. [doi: 10.1016/S0140-6736(11)60243-2] [Medline: 21684382]
4. Li G, Yin J, Gao J, Cheng TS, Pavlos NJ, Zhang C, et al. Subchondral bone in osteoarthritis: Insight into risk factors and microstructural changes. Arthritis Res Ther 2013;15(6):223 [FREE Full text] [doi: 10.1186/ar4405] [Medline: 24321104]
5. Madry H, Luyten FP, Facchini A. Biological aspects of early osteoarthritis. Knee Surg Sports Traumatol Arthrosc 2012 Mar;20(3):407-422. [doi: 10.1007/s00167-011-1705-8] [Medline: 22009557]
6. Cunha-Miranda L, Faustino A, Alves C, Vicente V, Barbosa S. Assessing the magnitude of osteoarthritis disadvantage on people's lives: The MOVES study [Article in Portuguese]. Rev Bras Reumatol 2015;55(1):22-30 [FREE Full text] [doi: 10.1016/j.rbr.2014.07.009] [Medline: 25582997]
7. Iversen MD, Price LL, von Heideken J, Harvey WF, Wang C. Physical examination findings and their relationship with performance-based function in adults with knee osteoarthritis. BMC Musculoskelet Disord 2016 Jul 12;17:273 [FREE Full text] [doi: 10.1186/s12891-016-1151-3] [Medline: 27405996]
8. Shakoor N, Foucher KC, Wimmer MA, Mikolaitis-Preuss RA, Fogg LF, Block JA. Asymmetries and relationships between dynamic loading, muscle strength, and proprioceptive acuity at the knees in symptomatic unilateral hip osteoarthritis. Arthritis Res Ther 2014 Dec 14;16(6):455 [FREE Full text] [doi: 10.1186/s13075-014-0455-7] [Medline: 25496937]
9. Joseph GB, Hilton JF, Jungmann PM, Lynch JA, Lane NE, Liu F, et al. Do persons with asymmetric hip pain or radiographic hip OA have worse pain and structure outcomes in the knee opposite the more affected hip? Data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2016 Mar;24(3):427-435 [FREE Full text] [doi: 10.1016/j.joca.2015.10.001] [Medline: 26497607]
10. Dobson F, Hinman RS, Roos EM, Abbott JH, Stratford P, Davis AM, et al. OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. Osteoarthritis Cartilage 2013 Aug;21(8):1042-1052 [FREE Full text] [doi: 10.1016/j.joca.2013.05.002] [Medline: 23680877]
11. Peter WF, Loos M, de Vet HC, Boers M, Harlaar J, Roorda LD, et al. Development and preliminary testing of a computerized animated activity questionnaire in patients with hip and knee osteoarthritis. Arthritis Care Res (Hoboken) 2015 Jan;67(1):32-39 [FREE Full text] [doi: 10.1002/acr.22386] [Medline: 24966068]
12. Zasadzka E, Borowicz AM, Roszak M, Pawlaczyk M. Assessment of the risk of falling with the use of timed up and go test in the elderly with lower extremity osteoarthritis. Clin Interv Aging 2015;10:1289-1298 [FREE Full text] [doi: 10.2147/CIA.S86001] [Medline: 26300633]
13. Almeida GJ, Schroeder CA, Gil AB, Fitzgerald GK, Piva SR. Interrater reliability and validity of the stair ascend/descend test in subjects with total knee arthroplasty. Arch Phys Med Rehabil 2010 Jun;91(6):932-938 [FREE Full text] [doi: 10.1016/j.apmr.2010.02.003] [Medline: 20510986]

14. Brooks GC, Vittinghoff E, Iyer S, Tandon D, Kuhar P, Madsen KA, et al. Accuracy and usability of a self-administered 6-Minute Walk Test smartphone application. Circ Heart Fail 2015 Sep;8(5):905-913 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.115.002062] [Medline: 26283292]
15. Samaan MA, Schwaiger BJ, Gallo MC, Link TM, Zhang AL, Majumdar S, et al. Abnormal joint moment distributions and functional performance during sit-to-stand in femoroacetabular impingement patients. PM R 2017 Jun;9(6):563-570. [doi: 10.1016/j.pmrj.2016.10.002] [Medline: 27729286]
16. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
17. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. J Clin Epidemiol 2012 Mar;65(3):301-308. [doi: 10.1016/j.jclinepi.2011.07.011] [Medline: 22169081]
18. Dobson F, Hinman RS, Hall M, Marshall CJ, Sayer T, Anderson C, et al. Reliability and measurement error of the Osteoarthritis Research Society International (OARSI) recommended performance-based tests of physical function in people with hip and knee osteoarthritis. Osteoarthritis Cartilage 2017 Nov;25(11):1792-1796 [FREE Full text] [doi: 10.1016/j.joca.2017.06.006] [Medline: 28647467]
19. Schwaiger BJ, Gersing AS, Lee S, Nardo L, Samaan MA, Souza RB, et al. Longitudinal assessment of MRI in hip osteoarthritis using SHOMRI and correlation with clinical progression. Semin Arthritis Rheum 2016 Jun;45(6):648-655 [FREE Full text] [doi: 10.1016/j.semarthrit.2016.04.001] [Medline: 27162009]
20. Hensor EMA, Dube B, Kingsbury SR, Tennant A, Conaghan PG. Toward a clinical definition of early osteoarthritis: Onset of patient-reported knee pain begins on stairs. Data from the osteoarthritis initiative. Arthritis Care Res (Hoboken) 2015 Jan;67(1):40-47 [FREE Full text] [doi: 10.1002/acr.22418] [Medline: 25074673]

Abbreviations

6MWT: 6-Minute Walk Test
ADL: activities of daily living
CR: repeatability coefficient
CST: Chair Stand Test
ICC: intraclass correlation coefficient
OARSI: Osteoarthritis Research Society International
PCC: Pearson's correlation coefficient
SCT: Stair Climb Test
SDtest-retest: standard deviation of test-retest means
SEM: standard error of measurement
UCSF: University of California San Francisco

Edited by G Eysenbach; submitted 06.08.17; peer-reviewed by N Dewan, T Jiang, H Umapathy; comments to author 24.08.17; revised version received 29.08.17; accepted 16.09.17; published 27.10.17.

Please cite as:
Adusumilli G, Joseph SE, Samaan MA, Schultz B, Popovic T, Souza RB, Majumdar S
iPhone Sensors in Tracking Outcome Variables of the 30-Second Chair Stand Test and Stair Climb Test to Evaluate Disability: Cross-Sectional Pilot Study
JMIR Mhealth Uhealth 2017;5(10):e166
URL: http://mhealth.jmir.org/2017/10/e166/
doi:10.2196/mhealth.8656
PMID:29079549

Original Paper

Mobile Technology Use Across Age Groups in Patients Eligible for Cardiac Rehabilitation: Survey Study

Robyn Gallagher[1], RN, PhD; Kellie Roach[2], RN; Leonie Sadler[3], RN; Helen Glinatsis[4], RN; Julie Belshaw[5], RN; Ann Kirkness[4], RN; Ling Zhang[1], RN; Patrick Gallagher[1], BSocSci; Glenn Paull[6], RN; Yan Gao[6], RN; Stephanie Ruth Partridge[7], BSNtr, PhD; Helen Parker[8], PhD; Lis Neubeck[9], RN, PhD

[1]Charles Perkins Centre, Sydney Nursing School, University of Sydney, Camperdown, Australia
[2]Ryde Hospital, Northern Sydney Local Health District, Sydney, Australia
[3]Manly Hospital, Northern Sydney Local Health District, Sydney, Australia
[4]Royal North Shore Hospital, Northern Sydney Local Health District, Sydney, Australia
[5]Hornsby Ku-ring-gai Hospital, Northern Sydney Local Health District, Sydney, Australia
[6]St George Hospital, South Eastern Sydney Local Health District, Sydney, Australia
[7]Charles Perkins Centre, Sydney School of Public Health, University of Sydney, Sydney, Australia
[8]Charles Perkins Centre, Faculty of Health Sciences, University of Sydney, Sydney, Australia
[9]School of Health and Social Care, Edinburgh Napier University, Edinburgh, United Kingdom

Corresponding Author:
Robyn Gallagher, RN, PhD
Sydney Nursing School
Charles Perkins Centre
University of Sydney
Building D17 City Road
University of Sydney
Camperdown, 2006
Australia
Phone: 61 0286270279
Fax: 61 0286272101
Email: robyn.gallagher@sydney.edu.au

Abstract

Background: Emerging evidence indicates mobile technology–based strategies may improve access to secondary prevention and reduce risk factors in cardiac patients. However, little is known about cardiac patients' use of mobile technology, particularly for health reasons and whether the usage varies across patient demographics.

Objective: This study aimed to describe cardiac patients' use of mobile technology and to determine variations between age groups after adjusting for education, employment, and confidence with using mobile technology.

Methods: Cardiac patients eligible for attending cardiac rehabilitation were recruited from 9 hospital and community sites across metropolitan and rural settings in New South Wales, Australia. Participants completed a survey on the use of mobile technology devices, features used, confidence with using mobile technology, willingness and interest in learning, and health-related use.

Results: The sample (N=282) had a mean age of 66.5 (standard deviation [SD] 10.6) years, 71.9% (203/282) were male, and 79.0% (223/282) lived in a metropolitan area. The most common diagnoses were percutaneous coronary intervention (33.3%, 94/282) and myocardial infarction (22.7%, 64/282). The majority (91.1%, 257/282) used at least one type of technology device, 70.9% (200/282) used mobile technology (mobile phone/tablet), and 31.9% (90/282) used all types. Technology was used by 54.6% (154/282) for health purposes, most often to access information on health conditions (41.4%, 117/282) and medications (34.8%, 98/282). Age had an important independent association with the use of mobile technology after adjusting for education, employment, and confidence. The youngest group (<56 years) was over 4 times more likely to use any mobile technology than the oldest (>69 years) age group (odds ratio [OR] 4.45, 95% CI 1.46-13.55), 5 times more likely to use mobile apps (OR 5.00, 95% CI 2.01-12.44), and 3 times more likely to use technology for health-related reasons (OR 3.31, 95% CI 1.34-8.18). Compared with the older group, the middle age group (56-69 years) was more than twice as likely to use any mobile technology (OR 2.42,

95% CI 1.27-4.59) and mobile technology for health-related purposes (OR 1.92, 95% CI 1.04-3.53). Participants who had completed high school were twice as likely to use mobile technology (OR 2.62, 95% CI 1.45-4.70), mobile apps (OR 2.05, 95% CI 1.09-3.84), and mobile technology for health-related reasons (OR 5.09, 95% CI 2.89-8.95) than those who had not completed high school. Associations were also present between participants living in metropolitan areas and mobile technology use (OR 1.07, 95% CI 1.07-4.24) and employment and mobile app use (OR 2.72, 95% CI 1.44-5.140).

Conclusions: Mobile technology offers an important opportunity to improve access to secondary prevention for cardiac patients, particularly when modified to suit subgroups. High levels of mobile technology use and health motivation need to be harnessed for secondary prevention.

(JMIR Mhealth Uhealth 2017;5(10):e161) doi:10.2196/mhealth.8352

KEYWORDS

mobile phone; cell phone; digital divide; cardiac rehabilitation; cardiovascular disease

Introduction

Cardiovascular disease (CVD) is a leading cause of death and disability globally [1]. Recurrence of cardiac events is common, causing frequent hospitalizations and high costs to the health system [2]. Secondary prevention is the key to limiting recurrence, yet patients struggle with initiating and maintaining the required behaviors [3]. An important evidence-based, cost-effective secondary prevention strategy is comprehensive cardiac rehabilitation (CR). Participation in CR reduces mortality and risk factors, as well as promotes recovery and quality of life [4,5]. Despite this, CR is underutilized, with less than one-third of eligible patients attending and dropout rates estimated at 25% [6]. A key factor contributing to poor CR participation is that delivery is in-person and offered at limited times and locations, so patients with limited resources, comorbidities, and other demands, such as caring roles, are unable to attend [7,8]. Technology, particularly, mobile devices that provide Internet access, offers a potential solution to reduce these barriers and improve access to secondary prevention strategies.

Advantages of mobile technologies for secondary prevention include timely patient education, real-time tracking of behavior, reminders, and prompts. Persuasive technology design and gaming principles can also be incorporated to promote key risk reduction across the life course [9,10]. Patients may also access health information and connect with health professionals and fellow cardiac patients more directly. Patients and health care providers may benefit from an increased capacity to compile, store, and deliver data, which may be used to assess and improve effectiveness. When mobile technologies are incorporated or offered as an alternative to traditional CR, improvements in multiple risk factors occur and mortality benefits have shown to be equal for both modes of delivery [11]. However, evidence regarding the benefits of specific mobile technology–based strategies for secondary prevention in cardiac patients is still evolving. Furthermore, implementation of these new strategies into practice is rare [12], in part because of lack of convincing evidence that cardiac patients are currently using mobile technology and the perceptions that the older age of this population will be a barrier.

Mobile Technology Use and Age

Mobile technologies have advanced rapidly and their adoption has been widespread in developed countries, with seniors showing the fastest adoption rates [13]. However, age is frequently perceived as a critical and a potential barrier to technology engagement because age influences the opportunities people have had to develop familiarity, skills, and confidence with technology from their education and employment experiences [14]. Barriers and facilitators may also be idiosyncratic to a particular technology or functionality for an age group [15]. People who are currently aged between 50 and 70 years tend to have used computers, Internet, email, and various other technologies and features in their work and daily life but, perhaps, not a mobile phone. When this age group does use a mobile phone, they tend not to use all the features, such as apps, or may not do so confidently [15]. Whereas, people aged under 50 years tend to have been exposed to multiple technologies through education and employment; therefore, they are more likely to confidently use the full extent of mobile phone features, including schedulers, apps, and social media. In contrast, people aged 70 years and older generally use devices in a more passive way, such as using a mobile phone for voice calls and receiving texts [16,17]. This older subgroup will turn to computers and tablet devices for Internet use [15], in part, for the bigger screen because of visual impairment, and they tend to rely on younger people in areas where they are less confident, such as for setup and problem solving [17]. Therefore, the influence of the patient's age is crucial to consider in any investigation of technology use for health [14].

Mobile Technology Use in Cardiac Rehabilitation Patients

Research into cardiac patients' engagement with mobile technology is in its infancy. Two studies were found that investigated technology use in CR patients of samples in New Zealand (n=74) [18] and Ireland and Belgium (n=298) [19]. The majority (97% [72/74] and 93.9% [280/298]) had a mobile phone, and mobile phone use was 38% (21/74) and 63.1% (188/298), respectively, with 74% (55/74) and 74.0% (220/298) of both samples accessing the Internet daily. Older patients were less likely to use mobile phone features or to be interested in Web-based CR programs [19]. The influence of education, employment, and confidence in mobile technology use was not assessed, and may have been significant, given that the samples from Ireland and Belgium were highly educated [19]. A more

thorough understanding of the use of technology devices and functionalities across age groups of cardiac patients is needed. This knowledge will ensure that health technology interventions can be developed with an understanding of the subgroup for whom they are most likely to benefit and or modified to ensure that the attributes and requirements are suitable to the larger population of cardiac patients.

This study aimed to describe cardiac patients' patterns of use of mobile technology and to determine the impact of age group after adjusting for education, employment, and confidence in mobile technology use.

Methods

Design and Patients

This multisite study involved a cross-sectional survey of cardiac patients, both in metropolitan settings (university [n=3] and community [n=3] hospitals) and rural settings (university [n=1] and community [n=2] hospitals), in New South Wales, Australia. Human research ethics approval was received from all institutions involved LNR/15 HAWKE/450.

Patients met inclusion criteria if they were: (1) current inpatients with a cardiac diagnosis and were eligible to be referred to CR, or (2) currently enrolled in a CR program, and (3) had sufficient understanding of the English language for consent and questionnaire processes. Patients with neurocognitive disorders and major visual impairment were excluded.

Sample size was calculated to be 250 patients, based on eight variables (gender, age group, home language English, education, marital/partner status, employment status, metropolitan or rural residence, confidence in mobile technology use) and on the basis of multiple regression analysis of technology engagement, and power was set at 80% and alpha=.05.

Current inpatients eligible for referral to and or patients currently attending CR were approached to participate in the study; once their consent was obtained, the survey was completed. Staff received training to ensure the survey process was standardized and remained present to assist if needed. A total of 296 patients were approached and 282 were recruited; reasons for refusal included not interested in being involved in research (n=6) and currently not using technology of any type and therefore not interested in this specific project (n=8).

Data Collection

Technology engagement was assessed using a 20-item survey combining components of questionnaires developed by Edwards et al [20] for use and confidence-in-use and Illiger et al [21] for use of mobile technology. All of the following questions were in checklist format with tick-box responses for when the item applied. Questions related to whether participants currently used technology devices (computer, tablet, mobile phone, voice/text only phone, and activity trackers) and features that were regularly used (voice calls, text messages, email, Internet, Skype/Facetime, mobile apps, social media, scheduling, and information access). Participants were then asked separately to identify the devices they, (1) felt confident in using; (2) could easily learn; and (3) would like to learn to use. Additionally, participants were then asked to identify any health-related use of the Internet to (1) access information on health and heart conditions, treatments, medications, and lifestyle change and (2) communication with health professionals or other heart patients.

Confidence with technology use was modified from the original questionnaires to refer to technology overall [20]. However, pilot testing of this item indicated that participants focused primarily on using new programs. Therefore, an item was created that assessed confidence with technology use based on how quickly participants felt they could use a new program on any device (1=very quickly to 4=very slowly). A pilot test of the full survey was conducted on 15 cardiac patients to assess the appropriateness of format and understanding of survey items, minor modifications were then made to improve readability, accuracy, and specificity.

Sociodemographic data (age, gender, ethnicity, home language, education level achieved, marital/partnership status, and employment) and clinical details were collected to characterize the sample and include in the analyses [22], Patients who indicated they did not use any technology completed the sociodemographic and clinical details only.

Statistical Analysis

Sociodemographic characteristics, engagement with different types of technology and functionalities were described using means, standard deviations (SD), frequencies, and percentages. Participants were grouped by age into categories of <56, 56-69, and >69 years to allow comparisons with the literature [14,16] and with reference to population level surveys of technology access and use [15,23]. The most relevant for the study context is the DeLoitte 2015 Australian technology survey, which categorized older Australians using a 68-year age threshold [15], and as the study recruited 1 year later than the report, 69 years of age was used. The final category was used to differentiate the age group for which technology was integral to their education and employment, in this case 56 years [24]. Comparisons between age groups were conducted using chi-square test for categorical variables and one-way analysis of variance followed by Tukey test for continuous variables. "Mobile technology" was defined as use of a mobile phone or tablet, and "health-related use" as Internet use to access health information or communicate with health professionals or other heart patients. The independent factors associated with mobile technology, mobile apps, and health-related Internet use were determined using simple linear regression analysis for each technology using the variables such as gender, age group, home language English, completed high school, marital/partner status, currently employed, metropolitan or rural resident, and overall confidence. All assumptions required for the linear regression analysis were met. The *P* value was set at .05 for all analyses, with Bonferroni correction to *P* value of .01 when multiple analyses occurred.

Results

The sample (n=282) had a mean age of 66.5 (SD 10.6, range 31-92) years, 72.0% (203/282) were male, and 79.1% (223/282)

lived in a metropolitan area (Table 1). All patients had at least one cardiac diagnosis, the most common being percutaneous coronary intervention (33.3%, 94/282), myocardial infarction (22.7%, 64/282), and coronary artery bypass graft surgery (22.3%, 63/282). The majority (91.1%, 258/282) of participants currently used at least one type of technology, 70.9% (200/282) used mobile technology (mobile phone/tablet), and 31.9% (90/282) used all types. The most common single technology used was desktop/laptop computers (68.1%, 192/282) followed by mobile phones (63.8%, 180/282), mobile phones were also the device reported most often reported as being used confidently, correspondingly 69.9% (197/282) and 62.8% (177/282) (Figure 1). Mobile phones and tablets were the types of technology that if not currently used, participants most often felt confident they could learn to use (41.1%, 116/282 and 37.9%, 107/282, respectively) and wanted to learn to use (13.1%, 37/282 and 14.2%, 40/282, respectively). As age increased, participants were less likely to use any mobile technology (mobile phone/tablet) (overall and post hoc bivariate analyses $P<.001$), and overall confidence for technology use decreased significantly (overall and post hoc bivariate analyses $P<.001$) (Table 1).

Table 1. Sample characteristics and technology use of study participants compared for age category (N=282).

Characteristics	Age category (years)			*P* value, across ages
	<56, N=44	56-69, N=123	>69, N=115	
Gender				
Male, n (%)	34 (77)	91 (74.0)	79 (68.7)	.52
Married/partnered, n (%)	33 (75)	82 (66.7)	69 (60.0)	.18
English primary language, n (%)	42 (95)	111 (90.2)	105 (91.3)	.79
Completed high school, n (%)	30 (68)	81 (65.9)	67 (58.3)	.36
Employed, n (%)	29 (65)	49 (39.8)	6 (5.2)	<.001
Metropolitan residence n (%)	35 (79)	96 (78.0)	94 (81.7)	.77
Technology use				
Mobile technology[a], n (%)	39 (88)	96 (78.0)	65 (56.5)	<.001[b]
Mobile apps, n (%)	31 (70)	52 (42.3)	24 (20.9)	<.001[b]
Health-related use, n (%)	32 (72)	74 (60.2)	49 (42.6)	.001[b]
Confidence (1-highest, 4-lowest), mean (SD)	2 (0.92)	2.41 (0.96)	2.67 (0.89)	.001[b]

[a]mobile phone or tablet.

[b]post hoc analyses all $P<.01$.

Figure 1. Use, confidence, and willingness to learn to use different technology devices.

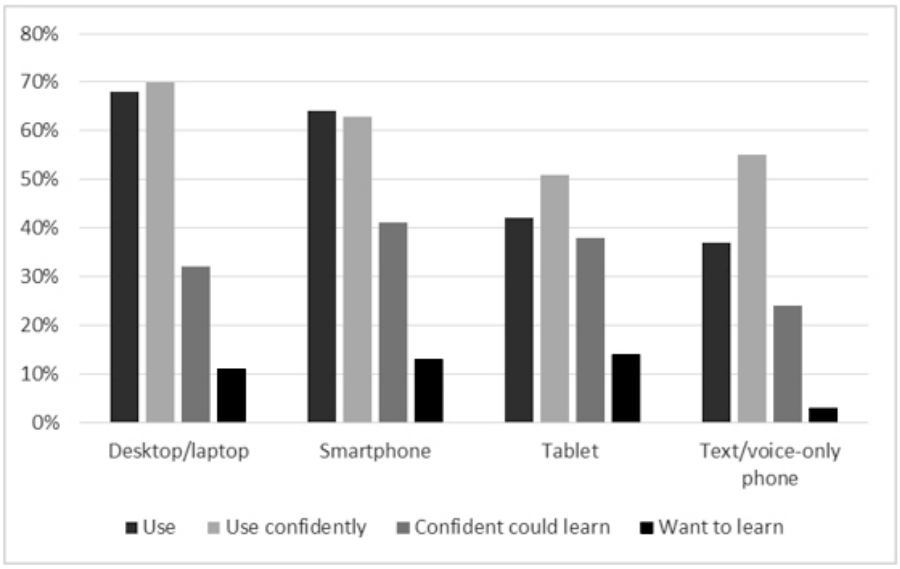
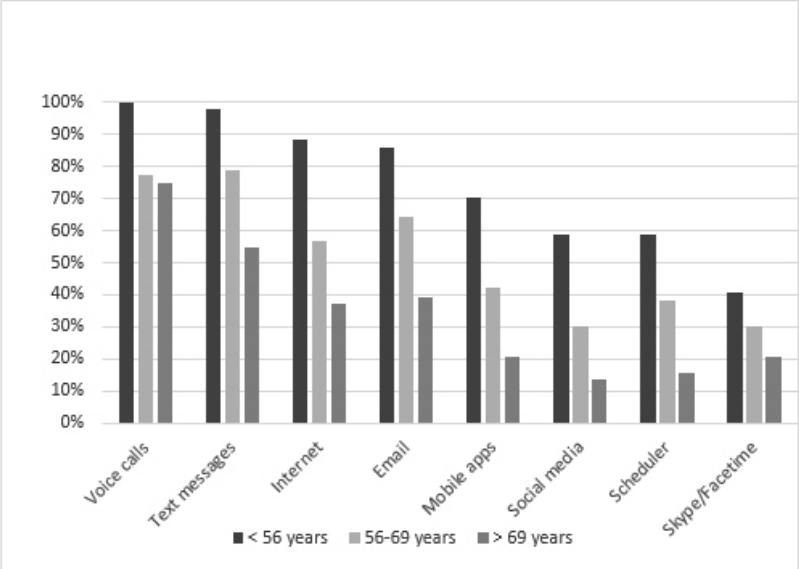


Figure 2. Use of technology features compared by age group.



Mobile technology (mobile phone/tablet) features used most often were voice calls (79.8%, 225/282), text messaging (70.6%, 199/282), sending email (55.3%, 156/282), using the Internet (52.1%, 147/282), and mobile apps (37.9%, 107/282). A small proportion (9.9%, 28/282) used all functionalities. With advancing age, every type of feature was used significantly less often, including mobile apps, with the exception of voice calls and Skype/Facetime (overall and post hoc bivariate analyses $P<.001$) (Figure 2,Table 1).

Technology was used by 54.6% (154/282) for health purposes, which included accessing health information, and this occurred most often for health conditions (41.5%, 117/282) and medications (34.8%, 98/282). As the age of the sample increased, health-related use decreased significantly (overall and post hoc bivariate analyses $P<.001$) (Table 1), including accessing information related to lifestyle changes, heart conditions and treatments, and communicating with health professionals (overall and post hoc bivariate analyses $P<.01$) (Figure 3). In contrast, accessing information on health conditions and medication information did not alter significantly. Different patterns of use were observed (not statistically tested) across age groups as the gap between mobile technology device use and the use of the device features, such as apps, was much larger in the two older groups (>69 years: mobile technology use of 56.5% [65/115] vs app use of 20.9% [24/115]; 56-69 years: mobile technology use of 78.0% [96/123] vs app use of 42.3% [52/123]), than in the youngest group (<56 years: mobile technology use of 88% [39/44] vs app use of 70% [31/44]) (Table 1). This gap was also present in health-related technology use but was much smaller and similar across age groups (>69 years: mobile technology use of 56.5% [65/115] vs health-related use of 42.6% [49/115]; 56-69 years: mobile technology use of 78.0% [96/123] vs health-related use of 60.2% [74/123]; and <56 years: mobile technology use of 88% [39/44] vs health-related use of 72% [32/44]).

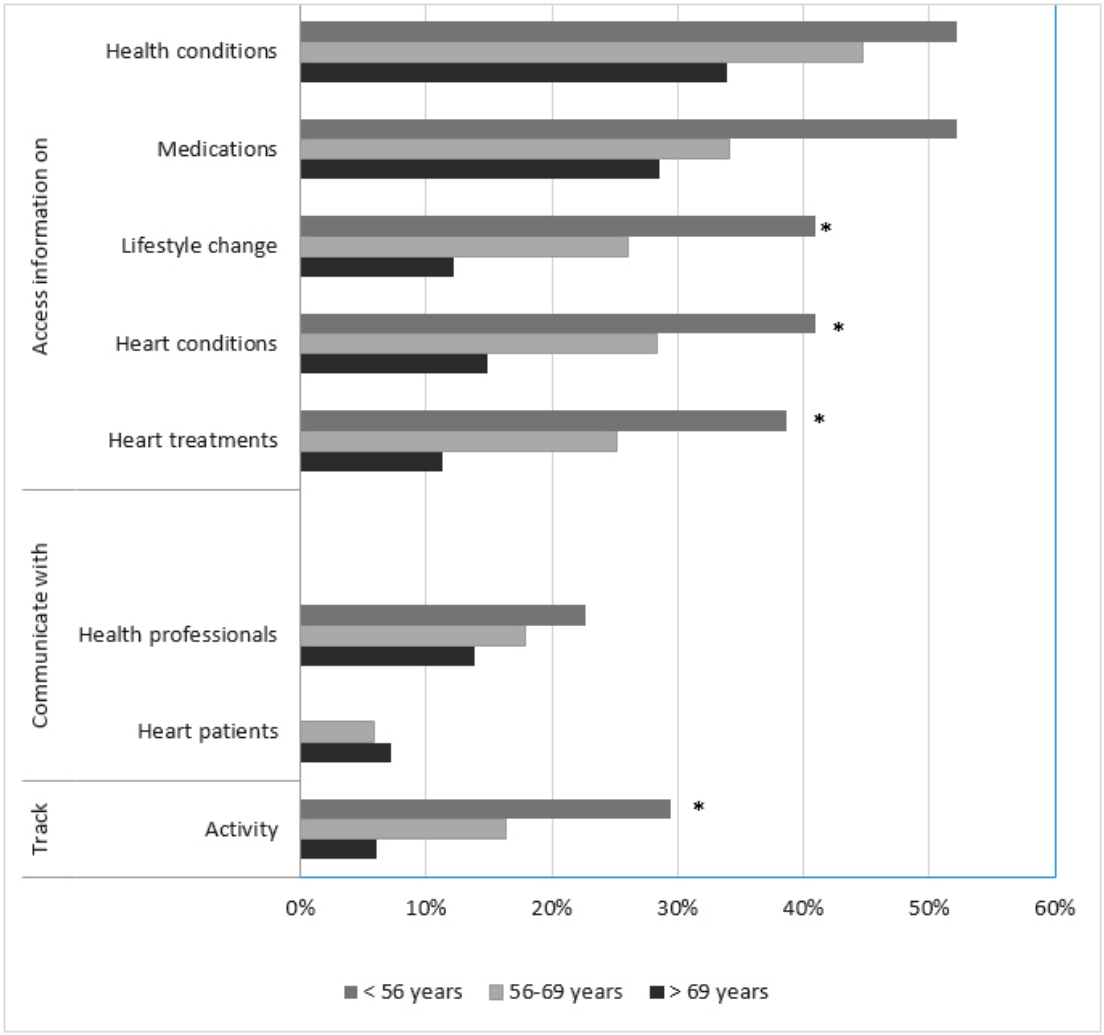
Age had an important independent association with mobile technology use after adjusting for education, and employment and other important variables (Table 2). Compared with the oldest age group, the youngest age group was at least four times more likely to use any mobile technology (odds ratio [OR] 4.45, 95% CI 1.46-13.55), 5 times more likely to use any mobile apps (OR 5.0, 95% CI 2.01-12.44), and 3 times more likely to use mobile technology for health-related reasons (OR 3.31, 95% CI 1.34-8.18). This association was evident but less pronounced when the middle age group was compared with the oldest age group, with participants more than twice as likely to use any mobile technology (OR 2.42, 95% CI 1.27-4.59) and mobile technology for health-related purposes (OR 1.92, 95% CI 1.04-3.53). Education was also important, with participants who had completed high school being much more likely to use any mobile technology (OR 2.62, 95% CI 1.45-4.70), mobile apps (OR 2.05, 95% CI 1.09-3.84), or to use mobile technology for health-related reasons (OR 5.09, 95% CI 2.89-8.95), rather than those who had not completed high school. Living in metropolitan areas increased the likelihood of any mobile technology use (OR 2.13, 95% CI 1.07-4.24), and employment increased the likelihood of using any apps (OR 2.72, 95% CI 1.44-5.14).

Table 2. Factors independently associated with mobile (mobile phone/tablet) technology use.

Characteristics	Mobile phone/tablet		Mobile apps		Health-related	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age						
<56 years vs >69 years	4.45 (1.46-13.55)	.009	5.00 (2.01-12.44)	.001	3.31 (1.34-8.18)	.01
56-69 years vs >69 years	2.42 (1.27-4.59)	.007	1.79 (0.93-3.45)	.08	1.92 (1.04-3.53)	.04
Completed high school	2.62 (1.45-4.70)	.007	2.05 (1.09-3.84)	.02	5.09 (2.89-8.95)	<.001
Metropolitan residence	2.13 (1.07-4.24)	.03	1.41 (0.65-3.04)	.39	1.63 (0.82-3.23)	.16
Employed	1.94 (0.87-4.33)	.12	2.72 (1.44-5.14)	.002	1.35 (0.69-2.59)	.39
Gender						
Male	1.66 (0.87-3.18)	.12	0.61 (0.32-1.14)	.12	0.67 (0.37-1.22)	.19

[a]OR: odds ratio.

Figure 3. Health-related technology uses compared by age group.



*Overall and post-hoc bivariate age group comparisons significant at *P*<.01

Discussion

Principal Findings

This study provides evidence that majority of patients eligible for or already attending CR, use mobile technologies such as mobile phones or tablets, providing the first evidence for the feasibility of using these technologies as an important alternative for delivering secondary prevention in more diverse samples. The study results contribute to the currently limited evidence that technology use is also very common when the sample is more diverse, including lower education and language backgrounds. Use of texting, Internet, and email were particularly high across all age groups. However, age and education were important influences in technology use and confidence of use. Younger and more educated patients were more likely to use mobile technology and to do so for health reasons, as well as to use apps, especially, if they were employed. Younger patients were also more confident in technology use.

Mobile technology use was high and comparable to the other limited studies in cardiac patients, despite the sample having much lower education levels. Mobile phone use (63.8%, 180/282), was similar to reports from Ireland and Belgium 63.1% (188/298), [19] but much higher than a sample from New Zealand 38% (21/74) [18]. An earlier recruitment year for the New Zealand sample may have contributed to this variation, despite being only 3 to 4 years, given the rapid penetration of mobile phones into the market and uptake of mobile technology in older groups such as cardiac patients [15]. Furthermore, this study identifies that the majority (54.6%, 154/282) of patients eligible for CR are using mobile technologies for health-related purposes [19], which was higher than reports from studies of general patient samples [16,21]. Common health-related uses included accessing health information, communicating with health professionals, and the use of activity-tracking devices [19]. It is important to capitalize on these health-related motivations given that a recent systematic review suggests that mobile health (mHealth) interventions can improve cardiovascular-related lifestyle behaviors and disease management in a way that is scalable to the public health level [25]. It is also important to acknowledge that cardiac patients use multiple sources of informational and behavioral support for their health and to include these aspects in patient education and recommend credible and trustworthy sources [9] would be helpful. Identification of any subgroups of users within cardiac patients and insight into associated differences within these subgroups of users is essential to the process.

Age was an important defining factor in cardiac patients' engagement with mobile technology, including for health-related purposes. This study adds to existing findings that a "digital divide" is present in mobile technology access and use for health reasons, and it also occurs in patients with cardiac conditions [14,16,26]. These previous studies proposed that the divide is the result of the relative presence of opportunities provided by education and employment that vary with age. This study is the first to identify that education and employment are indeed important, but the effect of age is also important and is independent of these aspects. Rather than a digital divide in mobile technology use created by age, education and employment, for cardiac patients the three age-defined groups identified by DeLoitte [15] proved accurate and reflected similarities to those identified by LeRouge et al [14]. For instance, younger individuals (<56 years) were highly engaged with mobile technology, using multiple devices and interactive features, such as apps, and frequently doing so for health. More than half of this group accessed online health and medication information, and more than a third accessed lifestyle and cardiac-related information and used trackers for their activity. The middle-aged group defined by LeRouge et al [14] as the Baby Boomer group (56-69 years) was also highly engaged with mobile technology, but their use was more narrowly focused, being far less likely to include interactive functionalities such as apps or trackers or to use mobile technology for health reasons. On the other hand, the oldest group (>69 years) were much less likely to be using mobile technology in all respects [14]. As a consequence, when mHealth interventions are developed, efforts should be made to ensure older patients are not accidentally excluded.

Aside from a lower likelihood of experience with mobile devices, aging is also accompanied by important changes in visual acuity and manual dexterity, which limits the potential use of small-screen devices [17]. Furthermore, the influence of older people's social group, including their peers, can limit motivation [27]. However, people of all ages can be taught to use technology, and so, although usage is less, benefits may still be obtained, particularly when larger screen devices, such as tablets, are used [17]. Therefore, age continues to be an important consideration in the development and targeting of mobile technology–based interventions for secondary prevention. However, this effect is likely to be rapidly diluted in the coming years, given the rapidly changing technology and communication landscape [15].

Education is another important factor to consider, potentially because it is an indicator of socioeconomic status and inequality generally. This study found that participants who had completed high school were at least five times more likely to be using mobile technology for health reasons than those not completing high school, which is consistent with a national survey on eHealth use in the United States [28]. In that survey of 2358 adult cancer patients, respondents who had not completed high school were less likely than even the lowest income participants to use the Internet/mobile phone for health reasons [28]. However, other idiosyncrasies in technology use were evident in that study, with the least educated and oldest males the least likely to engage in any mobile technology for health and younger females more likely to use social media for health. In this study, gender was not associated with mobile technology use in any respect. However, the Kontos et al [28] study highlights the need for a detailed understanding of the feasibility and acceptability of mobile technologies and features before developing online health-related interventions [29]. There is also an imperative to ensure when mobile technology is used for health that the information and support accessed is accurate and appropriate to their condition and circumstances. When these measures have been used for online secondary prevention

programs, uptake has been high and the intervention effective [30,31]. However, these examples are limited and further work is required, particularly to keep pace with rapidly changing technologies and features. Indeed, given the rapid adoption of technology by older people and the inevitable advancing age of the digital generation, there is an imperative for regular reassessment of health technology usage patterns [23].

Limitations

While the study recruited participants from multiple diverse locations, the sample may not represent all cardiac patients eligible for CR. The survey used to collect data was developed and modified based on previous studies and while pilot-tested in the relevant population, it has not been tested previously. The age categories chosen were based on relevant cut-points from the literature on technology use, resulting in unequal sample sizes, particularly for the youngest age group. If equal sized age groups were used, this would not reflect digital habits identified in the literature. The question used to assess overall confidence with technology requires further development to ensure a more comprehensive understanding of confidence with technology, particularly in relation to individual technology features.

Conclusions

This study identifies that mobile technology use in cardiac patients is at a high level, providing an important strategy for delivering secondary prevention, which should be harnessed. Furthermore, mobile technology offers an important opportunity to improve access to secondary prevention and enhance CR programs, particularly for younger patients for whom time and work pressures prove a barrier to participation. However, when developing mobile technology–based interventions, care must be taken not to presume that interventions demonstrated as applicable to younger age cardiac patients will also be directly applicable to older age patients.

Acknowledgments

This research was funded by the NSW Nursing and Midwifery Office Innovations Scholarship and seeding funds from the NSW Heart Foundation. The authors wish to thank Royal North Shore, Ryde, Manly, Hornsby Ku-Ring-Gai, Prince of Wales, St. George, John Hunter and Wagga Wagga Hospitals and Coffs Harbour Health Campus, and the staff and patients of the CR programs at these institutions, for their contributions to the questionnaire component of this study. In particular, the authors wish to thank Julie Belshaw, Ann Kirkness, Jenny Fildes, Dawn McIvor, Joanne Leonard, and Sheryl Bowen for their assistance with recruitment for this component of this study.

Conflicts of Interest

None declared.

References

1. Who.int/nmh/publications. Geneva: WHO Press; 2011 Apr. Global status report on noncommunicable diseases 2010 URL: http://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2017-10-14] [WebCite Cache ID 6uCuAzfCY]
2. Atkins ER, Geelhoed EA, Knuiman M, Briffa TG. One third of hospital costs for atherothrombotic disease are attributable to readmissions: a linked data analysis. BMC Health Serv Res 2014 Aug 8;14:338 [FREE Full text] [doi: 10.1186/1472-6963-14-338] [Medline: 25102911]
3. Shore S, Jones PG, Maddox TM, Bradley SM, Stolker JM, Arnold SV, et al. Longitudinal persistence with secondary prevention therapies relative to patient risk after myocardial infarction. Heart 2015 May 15;101(10):800-807 [FREE Full text] [doi: 10.1136/heartjnl-2014-306754] [Medline: 25801001]
4. Anderson L, Thompson DR, Oldridge N, Zwisler AD, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016 Jan 5(1):CD001800. [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
5. van Halewijn G, Deckers J, Tay HY, van Domburg R, Kotseva K, Wood D. Lessons from contemporary trials of cardiovascular prevention and rehabilitation: a systematic review and meta-analysis. Int J Cardiol 2017 Apr 1;232:294-303 [FREE Full text] [doi: 10.1016/j.ijcard.2016.12.125] [Medline: 28094128]
6. Suaya JA, Shepard DS, Normand ST, Ades PA, Prottas J, Stason WB. Use of cardiac rehabilitation by Medicare beneficiaries after myocardial infarction or coronary bypass surgery. Circulation 2007 Oct 9;116(15):1653-1662 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.701466] [Medline: 17893274]
7. Sun EY, Jadotte YT, Halperin W. Disparities in cardiac rehabilitation participation in the United States: a systematic review and meta-analysis. J Cardiopulm Rehabil Prev 2017 Jan;37(1):2-10. [doi: 10.1097/HCR.0000000000000203] [Medline: 27676464]
8. Soo Hoo SY, Gallagher R, Elliott D. Predictors of cardiac rehabilitation attendance following primary percutaneous coronary intervention for ST-elevation myocardial infarction in Australia. Nurs Health Sci 2016 Jun;18(2):230-237. [doi: 10.1111/nhs.12258] [Medline: 26858136]
9. Neubeck L, Lowres N, Benjamin EJ, Freedman SB, Coorey G, Redfern J. The mobile revolution--using smartphone apps to prevent cardiovascular disease. Nat Rev Cardiol 2015 Jun;12(6):350-360. [doi: 10.1038/nrcardio.2015.34] [Medline: 25801714]

10. Gallagher R, Neubeck L. How health technology helps promote cardiovascular health outcomes. Med J Aust 2016 Aug 1;205(3):107-108. [Medline: 27465761]
11. Neubeck L, Redfern J, Fernandez R, Briffa T, Bauman A, Freedman SB. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review. Eur J Cardiovasc Prev Rehabil 2009 Jun;16(3):281-289. [doi: 10.1097/HJR.0b013e32832a4e7a] [Medline: 19407659]
12. 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 Nov 1;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
13. Clark RA, Conway A, Poulsen V, Keech W, Tirimacco R, Tideman P. Alternative models of cardiac rehabilitation: a systematic review. Eur J Prev Cardiol 2015 Jan;22(1):35-74. [doi: 10.1177/2047487313501093] [Medline: 23943649]
14. LeRouge C, Van Slyke C, Seale D, Wright K. Baby boomers' adoption of consumer health technologies: survey on readiness and barriers. J Med Internet Res 2014 Sep 8;16(9):e200 [FREE Full text] [doi: 10.2196/jmir.3049] [Medline: 25199475]
15. Deloitte. 2015. Australian media and digital preferences - 4th edition URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/technology-media-telecommunications/deloitte-au-tmt-media-consumer-survey-2015-100815.pdf [accessed 2017-10-08] [WebCite Cache ID 6u4aLJPLG]
16. Hall AK, Bernhardt JM, Dodd V, Vollrath MW. The digital health divide: evaluating online health information access and use among older adults. Health Educ Behav 2015 Apr;42(2):202-209 [FREE Full text] [doi: 10.1177/1090198114547815] [Medline: 25156311]
17. Tsai HS, Shillair R, Cotten SR, Winstead V, Yost E. Getting grandma online: are tablets the answer for increasing digital inclusion for older adults in the U.S.? Educ Gerontol 2015;41(10):695-709. [doi: 10.1080/03601277.2015.1048165] [Medline: 26877583]
18. Dale LP, Whittaker R, Eyles H, Mhurchu CN, Ball K, Smith N, et al. Cardiovascular disease self-management: pilot testing of an mHealth healthy eating program. J Pers Med 2014 Mar;4(1):88-101 [FREE Full text] [doi: 10.3390/jpm4010088]
19. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. BMC Med Inform Decis Mak 2016;16:95 [FREE Full text] [doi: 10.1186/s12911-016-0329-9]
20. Edwards L, Thomas C, Gregory A, Yardley L, O'Cathain A, Montgomery AA, et al. Are people with chronic diseases interested in using telehealth? A cross-sectional postal survey. J Med Internet Res 2014 May 8;16(5):e123 [FREE Full text] [doi: 10.2196/jmir.3257] [Medline: 24811914]
21. Illiger K, Hupka M, von Jan U, Wichelhaus D, Albrecht UV. Mobile technologies: expectancy, usage, and acceptance of clinical staff and patients at a university medical center. JMIR Mhealth Uhealth 2014 Oct 21;2(4):e42 [FREE Full text] [doi: 10.2196/mhealth.3799] [Medline: 25338094]
22. Gallagher R, Belshaw J, Kirkness A, Warrington D, Sadler L, Roach K. Evaluation of a brief educational intervention to improve knowledge of sublingual nitroglycerin in cardiac rehabilitation patients. Eur J Cardiovasc Nurs 2013 Dec;12(6):529-535. [doi: 10.1177/1474515112473694] [Medline: 23315128]
23. Pewinternet. 2017. Americans' Internet Access: 2000-2015 URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [accessed 2017-10-08] [WebCite Cache ID 6u4aQBTeL]
24. Statista. 2014. Internet use by age group worldwide as of November 2014 URL: http://www.statista.com/statistics/272365/age-distibrution-of-internet-users-worldwide/ [accessed 2017-10-08] [WebCite Cache ID 6u4aWCAOj]
25. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015 Nov 24;132(21):2012-2027. [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
26. Bernhardt JM. Health education and the digital divide: building bridges and filling chasms. Health Educ Res 2000 Oct;15(5):527-531. [Medline: 11184212]
27. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to use mHealth in older adults With heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
28. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
29. Wiggers AM, Peek N, Kraaijenhagen R, Jaspers M. Determinants of eligibility and use of ehealth for cardiac rehabilitation patients: preliminary results. Stud Health Technol Inform 2014;205:818-822. [Medline: 25160301]
30. Varnfield M, Karunanithi MK, Särelä A, Garcia E, Fairfull A, Oldenburg BF, et al. Uptake of a technology-assisted home-care cardiac rehabilitation program. Med J Aust 2011 Feb 21;194(4):S15-S19. [Medline: 21401482]
31. 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 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]

Abbreviations

CR: cardiac rehabilitation

CVD: cardiovascular disease
eHealth: electronic health
mHealth: mobile health
OR: odds ratio
SD: standard deviation

Edited by G Eysenbach; submitted 05.07.17; peer-reviewed by M Varnfield, Y Tang; comments to author 17.08.17; revised version received 05.09.17; accepted 06.09.17; published 24.10.17.

Please cite as:
Gallagher R, Roach K, Sadler L, Glinatsis H, Belshaw J, Kirkness A, Zhang L, Gallagher P, Paull G, Gao Y, Partridge SR, Parker H, Neubeck L
Mobile Technology Use Across Age Groups in Patients Eligible for Cardiac Rehabilitation: Survey Study
JMIR Mhealth Uhealth 2017;5(10):e161
URL: http://mhealth.jmir.org/2017/10/e161/
doi:10.2196/mhealth.8352
PMID:29066425

Original Paper

A New Tool for Nutrition App Quality Evaluation (AQEL): Development, Validation, and Reliability Testing

Kristen Nicole DiFilippo[1], MS, RDN; Wenhao Huang[2], PhD; Karen M Chapman-Novakofski[3], PhD, RDN

[1]Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States

[2]Department of Education Policy, Organization and Leadership, University of Illinois at Urbana-Champaign, Champaign, IL, United States

[3]Department of Food Science and Human Nutrition, Division of Nutritional Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Corresponding Author:
Kristen Nicole DiFilippo, MS, RDN
Division of Nutritional Sciences
University of Illinois at Urbana-Champaign
238 Bevier Hall
905 S. Goodwin Ave
Urbana, IL, 61801
United States
Phone: 1 217 552 5777
Fax: 1 217 265 0925
Email: kdifilip@illinois.edu

Abstract

Background: The extensive availability and increasing use of mobile apps for nutrition-based health interventions makes evaluation of the quality of these apps crucial for integration of apps into nutritional counseling.

Objective: The goal of this research was the development, validation, and reliability testing of the app quality evaluation (AQEL) tool, an instrument for evaluating apps' educational quality and technical functionality.

Methods: Items for evaluating app quality were adapted from website evaluations, with additional items added to evaluate the specific characteristics of apps, resulting in 79 initial items. Expert panels of nutrition and technology professionals and app users reviewed items for face and content validation. After recommended revisions, nutrition experts completed a second AQEL review to ensure clarity. On the basis of 150 sets of responses using the revised AQEL, principal component analysis was completed, reducing AQEL into 5 factors that underwent reliability testing, including internal consistency, split-half reliability, test-retest reliability, and interrater reliability (IRR). Two additional modifiable constructs for evaluating apps based on the age and needs of the target audience as selected by the evaluator were also tested for construct reliability. IRR testing using intraclass correlations (ICC) with all 7 constructs was conducted, with 15 dietitians evaluating one app.

Results: Development and validation resulted in the 51-item AQEL. These were reduced to 25 items in 5 factors after principal component analysis, plus 9 modifiable items in two constructs that were not included in principal component analysis. Internal consistency and split-half reliability of the following constructs derived from principal components analysis was good (Cronbach alpha >.80, Spearman-Brown coefficient >.80): behavior change potential, support of knowledge acquisition, app function, and skill development. App purpose split half-reliability was .65. Test-retest reliability showed no significant change over time (P>.05) for all but skill development (P=.001). Construct reliability was good for items assessing age appropriateness of apps for children, teens, and a general audience. In addition, construct reliability was acceptable for assessing app appropriateness for various target audiences (Cronbach alpha >.70). For the 5 main factors, ICC (1,k) was >.80, with a P value of <.05. When 15 nutrition professionals evaluated one app, ICC (2,15) was .98, with a P value of <.001 for all 7 constructs when the modifiable items were specified for adults seeking weight loss support.

Conclusions: Our preliminary effort shows that AQEL is a valid, reliable instrument for evaluating nutrition apps' qualities for clinical interventions by nutrition clinicians, educators, and researchers. Further efforts in validating AQEL in various contexts are needed.

(JMIR Mhealth Uhealth 2017;5(10):e163) doi:10.2196/mhealth.7441

KEYWORDS

evaluation; mobile apps; dietitians; health education; diet, food, and nutrition

Introduction

Smartphone ownership reached 68% of Americans in 2015, increasing from 35% in 2011 [1]. Smartphones allow instant access to health information, enabling 62% of smartphone owners who obtain information on health conditions via smartphone [2]. A nationwide survey corroborated these results, showing 58% of mobile phone users in the United States had downloaded a health app, citing tracking physical activity (52.8%), tracking diet (46.6%), weight loss (46.8%), and to learn exercises (34.0%) as the most common reasons for health app use [3]. The study further suggested that research is needed to create methods to evaluate health app quality to ensure the needs of app users are met [3].

Dietitians are using apps in practice; a 2012 survey of Canadian dietitians showed 57.3% of dietitians surveyed used apps in practice, and 83.6% of those not currently using apps expressed interest in future app use in dietetic practice [4]. Whereas nutrition-related health apps are widely available and utilized, health professional's involvement in the development of apps' content and functionalities remains uncertain [5]. Currently, no method grounded in empirical studies for evaluating and selecting apps specifically for use in nutrition interventions exists. When selecting an app for dietetic practice, dietitians resort to subjectively relying on best clinical judgment or relying on similarly subjective recommendations of others [6].

Standardized app evaluation is called for to present cost-effective, transparent means of providing app developers and distributors with the necessary information to guide app selection [7]. A need for a systematic framework for evaluating health-based apps [8] and weight loss apps [9] have both been emphasized, although a recent investigation into best practices in health app evaluations emphasizes that an available best practice approach could not be identified [10]. The study did identify various constructs evaluated in studies, suggesting that a review of apps should include an evaluation of usability or functionality, a critique of potential to promote behavior change, and the quality of the health-related content [10]. None of the reviewed studies included an evaluation of all three constructs [10]. Whereas studies evaluating nutrition apps have reported the use of evidence-based treatment strategies [11], the use of theory [12], as well as behavior change techniques within apps [11,13-15], a measure of quality as perceived by the health care provider is needed, which evaluates the quality of the content and the functionality of the app to complement previous work evaluating scientific evidence that goes into mobile app development.

With this in mind, the objective of this study was the development, validation, and reliability testing of the app quality evaluation (AQEL) tool, an evaluation instrument for judging the quality of apps to aid in the development and selection of apps for nutrition interventions.

Methods

Survey Development

PubMed was searched first for nutrition and health education apps evaluation studies and then for educational website evaluation studies to form a pool of initial survey items. Search terms included website, app, and evaluation. The search was not limited by date. No apps evaluation studies were found, but 6 studies were identified that evaluated websites [13,16-20]. Three studies were excluded after review by 2 researchers. Two of these assessed only for the inclusion of behavior theories rather than broader measures of quality [13,19] and another did not provide specific questions within the paper [20]. The 3 evaluation tools selected to create the initial item pool were chosen based on relevance to education and coverage of items targeting the areas of content, usability, and technology [16-18].

Ninety-four items from these 3 selected website evaluation tools, with n=43 in one [16], n=16 in another [17], and n=35 in the third [18] were entered into a spreadsheet and sorted based on relevance to three categories: content, usability, and technology. One researcher completed the initial sorting of questions, with a second researcher reviewing the category selections. Any disagreement was discussed until agreement was reached. These categories were selected to broadly cover the needs of previously identified stakeholders in nutrition app development and use, namely the researchers and practitioners involved in content selection and distribution of the app as an educational tool [6], the end users (potential patients or clients) of the app in terms of usability [6,7], and developers of the app technology [7]. The 2 researchers removed items specific to websites with no relevance to apps, reworded other items to pertain to apps and nutrition, and divided complex questions into 2 or more questions. This resulted in 27 content, 9 technology, and 19 usability items. Additional items were created based on specific features of apps, including transition between pages and touchscreen functionality [21], with 6 new content items, 8 new technology items, 8 new usability items, and 2 questions identifying the app and the device used to download the app. As the three sources used different rating scales, questions were converted to 5-point Likert-type scales (content n=26, usability n=19, and technology n=13); yes or no, or yes, maybe, or no (content n=5, usability n=3, and technology n=1); or open ended questions (content n=2, usability n=5, and technology n=1) [15,22].

Content and Face Validation

Nutrition experts and app developers completed content validation by reviewing survey sections; app end users completed face validation. Institutional review board approval was obtained at all points where participants were involved. A total of 13 nutrition experts, including registered dietitians and nutrition professors with publications in app-based nutrition interventions, were contacted to review the 33 nutrition content questions, with 6 agreeing. Of 15 technology experts contacted, 4 agreed. For face validation, app users were recruited through

a Web-based weekly email newsletter at the University. This newsletter targets all university employees, not just academic faculty. The first 14 respondents were requested to review the 27 usability questions; 10 completed the review.

Each expert and app user was asked to review the survey selecting from the following options: complete the survey considering an app used in the past, complete the survey reviewing a new app, or provide general opinions of the survey questions. To specifically improve the validity of the survey, experts and app users were asked to cross out inappropriate questions, circle unclear words or phrases, modify unclear questions, add additional questions they felt would improve the survey, and provide any additional comments on survey items they felt would benefit survey development.

After modification based on expert panel review suggestions, further content and face validation was completed because of the magnitude of the changes and to allow review of the whole tool. Four of six nutrition expert panel reviewers repeated the procedures described above, reviewing all of 51 preliminary AQEL items.

Item Reduction

To reduce and evaluate the reliability of the preliminary AQEL items, nutrition professionals were recruited via an online discussion group from the Nutrition Education for the Public Dietetics Practice Group of the Academy of Nutrition and Dietetics. A total of 25 nutrition professionals evaluated 3 apps each using the 51 AQEL items. These apps were randomly assigned from a pool of 15 apps selected to represent a wide variety of nutrition-related apps, as described later on. This provided 75 evaluations using the preliminary AQEL items. The nutrition professionals completed a second evaluation of each app 3 weeks later, providing a total of 150 evaluations using the 51 preliminary AQEL items.

App selection specifically targeted 3 categories: popular apps, unpopular apps, and app-based games. Popular and unpopular apps were determined by searching the Apple App Store using 6 terms: healthy eating, nutrition, diet, nutrition games, diabetes, and diabetes recipes. This was completed daily (May 2014 to July 2014; January 2015 to February, 2015). In the App Store, the default setting was changed so that the apps were searched *by popularity*, and daily top apps were recorded. Nine apps ranked in the top 3 for their search term in both the 2014 and 2015 searches. Five of these were selected for reliability testing, including a calorie counter, a nutrition quiz, a digestive system game for kids, and 2 diabetes apps. An additional 4 apps were selected that were considered unpopular. Three had fallen in popularity, ranking in the top 6 of the 2014 search but not appearing in the 2015 search. These were a weight loss hypnosis app, a weight loss app, and a calorie tracker. One additional app was considered unpopular as it was the last English language app listed in a search of nutrition games on June 11, 2015. Six additional apps were selected to increase the number of educational gaming apps because of the specific interest of the research team to better understand educational games.

Item number was reduced first by removing 7 questions where not applicable was selected more than 50% of the time for the 150 evaluations, as the frequent selection of not applicable for a given item indicated that the question was considered by participants as irrelevant for evaluating apps in general. These included questions such as “how well does the app provide capacity to log food?” An additional 10 items were not included in principal component analyses (PCA) that allowed AQEL to be modified for the target audience of the app. Five of these items related to the specific age group the evaluator felt the app targeted, the other 5 to the specific educational needs of the app end user. For these items, the app evaluator selected the groups they would like to evaluate the app for; therefore, limiting the responses to these items.

The 34 remaining items were reduced into categories using PCA with varimax rotation with the 150 app evaluations. Items were removed and analysis rerun when communalities were less than .50. Factor criteria were Eigen values of 1 or more, at least two items per factor, primary loadings of .45 or more [23], and secondary loadings with a difference of at least .20. Additionally, only the number of factors required to explain just over 70% of the variance were retained. Scree plots were also examined for points of inflection to determine which factors to retain. For further refinement of factors, items not meeting these criteria were eliminated and additional factor analyses were run on the remaining factors with the factor number limited to the number of factors identified in the previous analysis.

Multiple imputation with 100 imputations followed by aggregation of imputations was used to treat missing data, with new imputations run each time items were removed.

Reliability Analysis

Construct reliability of the final factors was assessed using Cronbach alpha. Spearman-Brown coefficient was used to test split-half reliability. For construct reliability only, items not on a 5-point scale were adjusted to a 5-point scale. These analyses were conducted using the first occasion apps were evaluated. Each rater’s first evaluation was used for analysis (n=75).

Items within each factor were summed to create factor scores. Test-retest analysis was conducted comparing first and second evaluations using Wilcoxon sign-rank as the data were not normally distributed (n=75).

Interrater reliability (IRR) for the evaluation of each app using factors identified in PCA was tested using one-way random, average measures intra correlations (ICC) using the first evaluation (n=75).

For the items assessing app appropriateness for various age groups (n=5) and target audiences (n=5), construct reliability was measured using Cronbach alpha. For the questions regarding age group, the second evaluation completed by each evaluator was utilized because of a mistake in the questionnaire discovered after many of the first evaluations had been completed. For the target audience, the first evaluation of the app by each evaluator was used. Sample size varied, as evaluators were able to select the age groups and target audiences. All sample sizes are reported.

Further IRR testing of app evaluations using the factors identified in PCA plus the age and audience constructs utilized

two-way random, average measures ICC. For this analysis, a new dataset was collected, with 15 nutrition professionals using the AQEL tool to evaluate MyFitnessPal, the most popular app according to a dietitians' survey (unpublished data, 2017) [24]. For this analysis, the age group apps evaluated for was *adults* and the evaluators considered the target audience of *people seeking weight loss support* (n=15).

Reporting on the survey using the Checklist for Reporting of Internet E-Surveys can be found in Multimedia Appendix 1 [25]. All statistical analysis were conducted using the Statistical Package for the Social Sciences (SPSS) for Windows, version 24 (IBM Corp).

Results

Content and Face Validation

Specific recommendations from the nutrition experts included clarifying words such as aim and target population, changing the rating scales used, and requesting additional items on skill-building and goals of the apps. App users recommended reducing repetitive questions. Technology experts recommended clarification of 10 items and dividing 3 items into multiple questions plus additional items on data storage and user characteristics. The recommendations of the expert panels led to the modification of nearly every item. Once modifications were completed as described, the three subtools were combined into the full AQEL with 51-scaled items for evaluating app quality plus items for app identification. The second expert panel resulted in minor clarifications.

Item Reduction

For the first round of PCA, Kaiser-Meyer-Olkin measure of sampling adequacy was .59 and the Bartlett test of sphericity was significant (χ^2_{561}=4456, *P*<.001). Correlations were greater than .30, and communalities were greater than .50 for all items. Nine factors had Eigen values greater than 1, but 8 factors explained 73% of the variance. When using 9 factors, 3 items were removed because they loaded onto 2 factors with less than a .20 difference; 1 item was removed because it did not load at .45 on any factor. Removing these items resulted in the elimination of a factor; therefore, the next analysis was run with 8 factors.

In the second PCA analysis, Kaiser-Meyer Olkin measure of sampling adequacy was .67, Bartlett test of sphericity was significant (χ^2_{435}=3357, *P*<.001). The point of inflection on the scree plot indicated that 5 factors should be retained (Figure 1); therefore, analysis was rerun with 5 factors. Five items had communalities below .50; these item were removed, and PCA was completed a third time with 5 factors.

In the final PCA analysis, Kaiser-Meyer-Olkin measure of sampling adequacy was .81, and Bartlett test of sphericity was significant (χ^2_{300}=2929, *P*<.001). All correlations were greater than .30, and communalities were .50 or greater when rounded to the nearest tenth. For items that loaded on more than one factor, differences were greater than .20 when rounded to the nearest tenth, and these items were placed on the factor where they loaded the highest. The final factor loadings are presented in Table 1.

Figure 1. Scree plot for second round of principal components analysis of items assessing nutrition app quality.

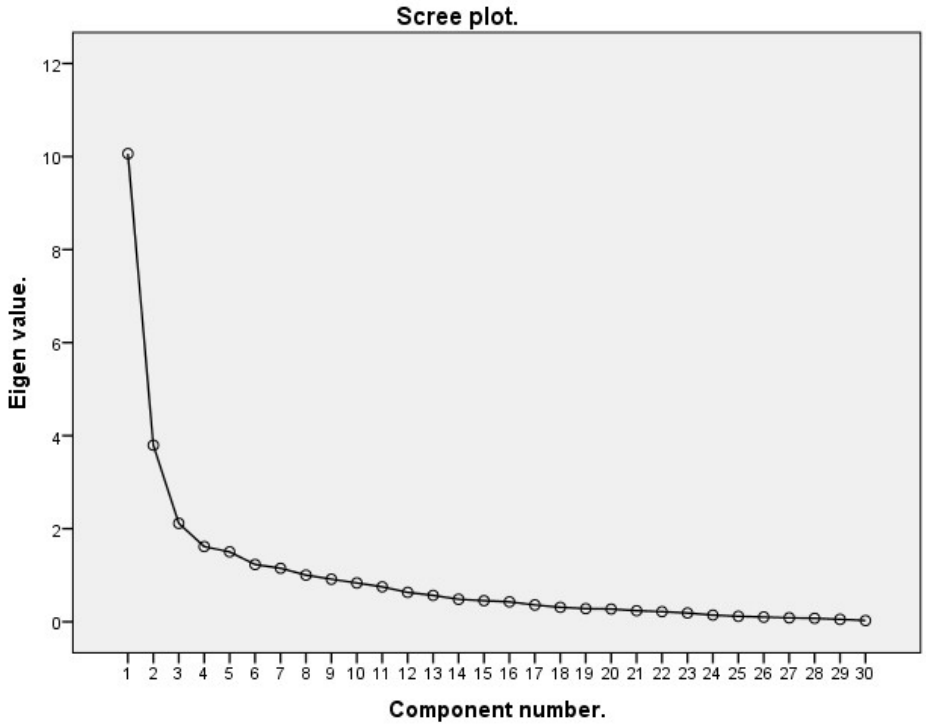


Table 1. Primary factor loadings of items assessing nutrition app quality.

Item	Factor loading value
Factor 1-Behavior change potential	
In your opinion does the app try to change behavior?	.56
Do you think the app will lead to behavior change?	.81
When considering activities within the app, will the activities help the user to change behavior?	.80
Would your friends use this app?	.57
Do you intend to use this app in the future?	.59
Will you do something differently after using this app?	.82
Will you try to do something new after using this app?	.82
Factor 2-Support of knowledge acquisition	
In your opinion, does the app try to increase knowledge?	.78
Do you think the app will increase the user's knowledge?	.69
When considering activities within the app, will the activities help the user to increase knowledge?	.70
How well does the app provide information?	.71
How well does the app provide feedback on progress?	.75
How well does the app provide timely feedback whenever needed?	.77
Is feedback provided when the user participates in an activity in the app?	.67
Factor 3-App function	
Please rate the speed of loading the app	.62
Please rate the user's ability to retrace their steps if they need to	.80
Please rate the transitions from page to page	.87
Please rate the function of any animations (quick and functional—slow and fragmented)	.60
Please rate the design of menus and icons	.79
Please rate the ease of navigation to the app's various features	.81
Factor 4-Skill development	
In your opinion does the app try to develop a skill?	.89
Do you think the app will lead to the development of a skill?	.71
When considering activities within the app, will the activities help the user to develop a skill?	.70
Factor 5-App purpose	
Do you feel that the app has a clear purpose?	.68
Does the app title accurately describe the content of the app?	.76

Missing data represented 6.40% (240/3750) of the entries in the dataset. Little missing completely at random (MCAR) test indicated data were not MCAR (χ^2_{1140}=1267, *P*=.005); therefore, multiple imputations with 100 imputations were used to treat missing data.

Reliability Analysis

Measures of reliability are reported in Table 2. Construct reliability for factors 1 to 4 were all excellent with Cronbach alpha between .80 and .90 [26]. Split-half reliability for factors 1 to 4 were also good with Spearman-Brown coefficients between .80 and .90. For factor 5, Cronbach alpha was not used as there were only two items, and split half reliability was .65.

Test-retest reliability was not significant, indicating that evaluations of each factor did not change over time, with the exception of the factor evaluating the potential of the app assist skill development (Table 3).

IRR for each app was excellent [27]. These results are shown in Table 4. The 15th app was not included, as only one evaluator completed the evaluation for this app.

For the 5 items assessing specific age groups, construct reliability was good except for evaluations specific to adults (Table 5) [26].

For the 5 items assessing the app's appropriateness of various audience, construct reliability was less than desirable; however, removing one item improved the reliability to be at minimum acceptable (>.70) and for many audiences good (>.80) or

excellent (>.90) [19,26]. These results are presented in Table 6.

In the second dataset with 15 nutrition professionals evaluating My Fitness Pal, two-way random ICC using average measures was excellent, with ICC (2,15)=.99, *P*<.001. Single measure ICC was also good, with ICC (2,15)=.83, *P*<.001, reflecting that the AQEL tool, including both the factors identified by PCA and the two additional modifiable constructs, can be reliably used both by averaging responses of multiple evaluators or by a single evaluator.

Table 2. Construct and split-half reliability of factors evaluating app quality (n=75 evaluations).

Factor	Construct reliability (Cronbach alpha)	Split-half reliability (Spearman-Brown coefficient)
Factor 1-Behavior change potential	.89	.82
Factor 2-Knowledge	.88	.84
Factor 3-App function	.89	.83
Factor 4-Skill development	.81	.83
Factor 5-App purpose	N/A[a]	.65

[a]Cronbach alpha not applicable as the factor only includes 2 items.

Table 3. Test-retest reliability of factors of the app quality evaluation (n=75 evaluations).

Factor or item	Test-retest reliability Wilcoxon signed-rank test (*P* value)
Factor 1-Behavior change potential	.13
Factor 2-Knowledge	.05
Factor 3-App function	.55
Factor 4-Skill development[a]	.001
Skill item 1	.01
Skill item 2	.006
Skill item 3	.05
Factor 5-App purpose	.89

[a]Results for individual items shown for skill development as differences were found to be significant.

Table 4. Interrater reliability of dietitians evaluating apps using the app quality evaluation (AQEL) tool (n=75 evaluations).

App	ICC[a](1,k)	*P* value
Calorie Counter by MyFitnessPal	ICC (1,3)=.88	.003
Nutrition Quiz 600+ Facts, Myths and Diet Tips	ICC (1,3)=.94	<.001
Science Heroes: Digestive System for Kids by Yogome Inc.	ICC (1,5)=.86	.001
Diabetes In Check: Coach, Blood Glucose & Carb Tracer by Everyday Health Inc.	ICC (1,8)=.96	<.001
Diabetes App Lite by BHI Technologies, Inc.	ICC (1,6)=.96	<.001
Weight Loss Hypnosis-Free by Surf City Apps LLC	ICC 1,4)=.80	.01
Jillian Michael's Slim Down	ICC (1,4)=.87	.002
MyPlate Calorie Tracker	ICC 1,5)=.96	<.001
National Center on Health Nutrition Education Gamelettes by ZebraZapps Engineering	ICC (1,5)=.97	<.001
Nutrition and Healthy Eating by Tribal Nova	ICC (1,7)=.95	<.001
Awesome EatsTM by whole Kids Foundation	ICC (1,6)=.95	<.001
Eat Smart by Edin	ICC (1,7)=.92	<.001
Eat & Move O-Matic by Learning Games Lab, NM State University	ICC (1,6)=.96	<.001
Harry's Healthy Garden	ICC (1,5)=.98	<.001

[a]ICC: intraclass correlations.

Table 5. Construct reliability of items assessing app appropriateness for evaluator-selected age groups.

Age group	Evaluations completed (n)[a]	Construct reliability (Cronbach alpha)
Children	36	.82
Teens	12	.86
Adults	31	.53
General audience	10	.80
Other audience	6	N/A[b]

[a]Analysis of responses from evaluators second evaluation of each app because of survey error discovered during first round of evaluations.
[b]Analysis not completed because of negative covariance among items.

Table 6. Construct reliability of items assessing app appropriateness for evaluator-selected audiences.

Audience	Evaluations completed (n)	Construct reliability (Cronbach alpha)	Construct reliability with item 5 deleted[a] (Cronbach alpha)
People seeking help for medical conditions	16	.62	.82
People with specific nutrition concerns	5	.67	.94
People who are shopping for food	3	.40	.98
People seeking recipe or meal ideas	8	.20	.70
People seeking guidance for restaurant eating[b]	1	-	-
People seeking weight loss support	18	.53	.92
People seeking nutrition education	43	.57	.71
Other audience	16	.59	.72

[a]Item 5: Does the level of detail exceed the target populations' abilities?
[b]Analysis not run as only 1 person selected this option.

Discussion

Principal Findings

In summary, the 94 items first selected from the literature were modified to 51 items after expert panel review. Five items evaluating app appropriateness for various age groups and 5 items evaluating app appropriateness for evaluator chosen target audiences were not included in PCA, as the response number to these items was limited. Construct reliability testing of these two constructs resulted in removal of one item evaluating appropriateness for target audiences. This left 41 items to be grouped into factors for evaluating apps. Seven of these were eliminated as raters selected the option of does not apply in more than 50% of the evaluations. Therefore, 34 items were tested using PCA. After three rounds of PCA, the result was a survey with 25 items grouped into 5 factors for evaluating apps, plus 5 additional items that can be used for evaluating app appropriateness for various age groups, and 4 additional items which can be used to evaluate apps for specific target audiences (Multimedia Appendix 2).

The AQEL is a valid, reliable tool for evaluating app quality. Careful consideration of stakeholder needs, including nutrition educators and researchers, app end users, and developers, guided development and assurance of face and content validity. Construct, split-half, test-retest, and IRR were also evaluated to establish the overall reliability of this new tool for use in evaluating nutrition apps.

The validation and reliability testing of AQEL contributes to the literature by providing a standardized method of evaluating and reporting on nutrition apps, a gap identified previously in app research [7-10].

Limitations

AQEL allows for the evaluator to specifically choose both an age group and audience for some of the evaluation items. Although a strength of the tool, it did limit the samples size for reliability testing of these items. Addressing characteristics of the intended app user, such as learning preferences and skill with technology, is an important aspect of selecting apps. Generally apps are able to accommodate a wide variety of user preferences as they are able to deliver multimedia content based on users' choices. Rarely do apps only deliver text or multimedia content. Assessment based on age group begins to address variations in app users; however, clinician assessment remains an important piece when selecting apps for clients to account for individual preferences and needs.

App users for face validation were recruited from a population of university employees reflecting a wide range educational experience and income; however, demographic data were not collected from this group.

Rater knowledge of apps is important for completing accurate evaluations. For this reason, raters were asked to spend 10 to 15 min becoming familiar with the app before completing the evaluation [28]. Whereas extensive repeated use of the app would be ideal, this is not always feasible in practice, especially when evaluating a large number of apps. Test-retest reliability showed that for most questions, results remained stable as raters presumably were more familiar with apps on the second evaluation compared with the first; however, raters were not asked how familiar they were with the app on the first evaluation. Future studies comparing AQEL ratings on first use with later evaluations after regular use of an app would be useful to corroborate this finding.

During validation, a mistake was discovered in the display logic of Q17 to Q21 in the survey. These questions all concerned the subscore of the category appropriateness to the target audience. Of the included 75 surveys 37 had been completed at the time the error was identified. The survey was corrected and updated. To account for this, the analysis of Cronbach alpha for items evaluating app appropriateness for selected age groups were taken from the second evaluation of each app.

Comparison With Prior Work

Previous studies evaluating nutrition apps focus primarily on evidence-based features currently available in apps [11,12] and behavior change techniques or behavior theory use within apps [11,13-15]. AQEL provides the first valid and reliable instrument specifically for dietitians and nutrition researchers to evaluate the quality of apps for use in nutrition interventions. AQEL would add to such evaluations by providing a quantitative method of scoring app quality.

One app selection method in chronic disease management calls for practitioners to create an app library by identifying apps per topic. Evaluation of these apps are based on popularity and incorporation of best practice guidelines, assessing the use of behavior theory using the Behavioral Theory Content Survey [13], then matching apps to patient preferences and disease etiology [29]. This methodology, inevitably, depends heavily on the popularity rating of an app and requires an individual subjective judgment of the quality of many apps. AQEL could add to this methodology by proving an objective measure of app quality specific to content for nutrition education.

When evaluating apps in research, it has been recommended to evaluate apps in terms of what works, for which people, and in what circumstances [6]. AQEL allows for this by considering not just the app but also the end user. AQEL is consistent with a previous study evaluating platforms supporting apps, incorporating the same perspectives of developer, end user, and content provider [30].

At the onset of this study, no tools had been developed for app evaluation. During development of the AQEL tool, another app evaluation tool, the mobile app rating scale (MARS) was published for health app evaluation [28]. This 23-item tool included 5 subscales for measuring app quality: engagement, functionality, aesthetics, information, and app subjective quality. MARS also supplies optional items that can be modified to assess knowledge, attitudes, and intention to change; however, these are not included in the main scoring of MARS. Reliability testing for MARS was completed using evaluation of mental health apps with overall two-way mixed ICC=.79, 95% CI 0.75-0.83, whereas the subscales ICC=.50 to .80. Cronbach alpha of subscales=.80 to .89, median=.85, and overall scale=.90. Although not originally designed for nutrition apps, one recent study used the first four MARS categories to evaluate weight loss apps, finding that IRR between 2 raters was good, with median Krippendorf alpha=.80 and interquartile range=.14 [31].

However, AQEL differs from MARS in several important ways. First, development and reliability testing of AQEL was specifically based on input from practicing and research nutrition professionals. ICC testing in MARS relied on 2 raters' evaluations, whereas reliability testing of the AQEL used 25 raters. PCA was used to refine AQEL, and test-retest reliability was evaluated; steps not included in the testing of MARS. Second, AQEL includes as primary constructs the categories of behavior change potential, knowledge, and skill development. These categories are not captured as part of the main MARS score; instead, there are optional items assessing similar categories: knowledge, attitudes, and intentions to change. Reliability testing of these categories is not provided for MARS. Behavior change potential, which is included in AQEL but not MARS, along with functionality and the appropriateness and quality of content for the targeted health condition, which are included in both scales, have been cited as critical for a complete evaluation of health-related apps [10]. MARS and AQEL both allow for modification of items concerning the targeted health behavior or audience; only AQEL allows for modification of items based on the targeted age group being considered. This allows greater flexibility as a dietitian could rate the same app differently when considering two different age groups. Finally, AQEL places a clear emphasis on evaluating the ability of the app to support education to increase nutrition knowledge and support behavior change.

An additional checklist was recently published for physician use in evaluating health apps; however, no information is provided on development, validation or reliability testing, and no scoring scheme was provided [32].

Conclusions

The AQEL is a reliable tool for use when designing educational interventions that include nutrition-related apps. This tool fills a gap by allowing for standardized evaluation of the vast number of apps available for use in dietetics practice and research that have not undergone rigorous testing [6,9]. By providing evaluation based on multiple factors of quality, app selection can focus on the specific needs of the client. For example, if looking for an app specifically to support behavior change, those scores can be focused on, while also evaluating for functionality and appropriateness for the age and nutrition needs of the client. Scores from the scale can be evidence to justify app selection for interventions as well. Additionally, this tool will help inform app selection in future studies assessing for consistent use, behavior change, and improved clinical outcomes, and to provide dietitians with standardized reports [8-10] on the strengths and weaknesses of apps available to their clients.

Acknowledgments

This work was supported by the USDA National Institute of Food and Agriculture Hatch project through the Division of Nutritional Sciences Vision 20/20 program at the University of Illinois, Hatch Project # ILLU-9710362.

The authors would like to thank Steve Zazeski and Weixiao Huang for technical assistance in identifying apps for this project.

Authors' Contributions

KD was responsible for survey development, participant recruitment, data collection, statistical analysis, and manuscript preparation. WH was responsible for study design, interpretation of results, and manuscript preparation. KCN was responsible for institutional review board's approval, study design, analysis, and manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 16KB - mhealth_v5i10e163_app1.pdf]

Multimedia Appendix 2

App quality evaluation.

[PDF File (Adobe PDF File), 82KB - mhealth_v5i10e163_app2.pdf]

References

1. Anderson M. Technology device ownership. 2015. 2015 URL: http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/ [accessed 2017-02-02] [WebCite Cache ID 6nyjYHNno]
2. Smith A, McGeeney K, Duggan M, Rainie L, Keeter S. Pewinternet. 2015. U.S. smartphone use in 2015 URL: http://www.pewinternet.org/files/2015/03/PI_Smartphones_0401151.pdf [accessed 2017-02-01] [WebCite Cache ID 6nyjpzFCv]
3. 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]
4. Lieffers JRL, Vance VA, Hanning RM. Use of mobile device applications in Canadian dietetic practice. Can J Diet Pract Res 2014;75(1):41-47. [Medline: 24606959]
5. Harricharan M, Gemen R, Celemín LF, Fletcher D, de Looy AE, Wills J, et al. Integrating mobile technology with routine dietetic practice: the case of myPace for weight management. Proc Nutr Soc 2015 May;74(2):125-129. [doi: 10.1017/S0029665115000105] [Medline: 25804507]
6. Hingle M, Patrick H. There are thousands of apps for that: navigating mobile technology for nutrition education and behavior. J Nutr Educ Behav 2016 Mar;48(3):213-8.e1. [doi: 10.1016/j.jneb.2015.12.009] [Medline: 26965099]
7. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
8. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Feb;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
9. Chen J, Cade JE, Allman-Farinelli M. The most popular smartphone apps for weight loss: a quality assessment. JMIR mHealth uHealth 2015 Dec 16;3(4):e104. [doi: 10.2196/mhealth.4334] [Medline: 26678569]
10. 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 2016 Oct 18:pii: 1357633X16673538 [Epub ahead of print]. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
11. Dute DJ, Bemelmans WJ, Breda J. Using mobile apps to promote a healthy lifestyle among adolescents and students: a review of the theoretical basis and lessons learned. JMIR Mhealth Uhealth 2016 May 05;4(2):e39 [FREE Full text] [doi: 10.2196/mhealth.3559] [Medline: 27150850]
12. Juarascio AS, Manasse SM, Goldstein SP, Forman EM, Butryn ML. Review of smartphone applications for the treatment of eating disorders. Eur Eat Disord Rev 2015 Jan;23(1):1-11 [FREE Full text] [doi: 10.1002/erv.2327] [Medline: 25303148]
13. 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]

14. 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]
15. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
16. Alyusuf RH, Prasad K, Abdel Satir AM, Abalkhail AA, Arora RK. Development and validation of a tool to evaluate the quality of medical education websites in pathology. J Pathol Inform 2013;4:29 [FREE Full text] [doi: 10.4103/2153-3539.120729] [Medline: 24392243]
17. Joshi MP, Bhangoo RS, Kumar K. Quality of nutrition related information on the internet for osteoporosis patients: a critical review. Technol Health Care 2011;19(6):391-400. [doi: 10.3233/THC-2011-0643] [Medline: 22129940]
18. Minervation. Full LIDA tool URL: http://www.minervation.com/lida-tool/minervation-lida-instrument-v1-2/ [accessed 2017-02-02] [WebCite Cache ID 6nyjw2QoJ]
19. Doshi A, Patrick K, Sallis JF, Calfas K. Evaluation of physical activity web sites for use of behavior change theories. Ann Behav Med 2003;25(2):105-111. [Medline: 12704012]
20. Guardiola-Wanden-Berghe R, Gil-Pérez JD, Sanz-Valero J, Wanden-Berghe C. Evaluating the quality of websites relating to diet and eating disorders. Health Info Libr J 2011 Dec;28(4):294-301. [doi: 10.1111/j.1471-1842.2011.00961.x] [Medline: 22051128]
21. Drew J. Journalofaccountancy. 2013. How to develop and publish a mobile app URL: http://www.journalofaccountancy.com/issues/2013/feb/20126350.html [accessed 2017-02-02] [WebCite Cache ID 6nywylEvP]
22. Russ-Eft D, Preskill H. Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change. 2nd ed. New York, NY: Basic Books; 2009.
23. Hair Jr JF, Tatham RL, Anderson RE, Black W. Multivariate Data Analysis. 5th ed. London: Prentice-Hall; 1998:0138948585.
24. Karduck J, Chapman-Novaksofski K. Results of the Clinician Apps Survey, how clinicians working with patients with diabetes and obesity use mobile health apps. J Nutr Educ Behav 2017 (forthcoming).
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. George D, Mallery P. SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 update. Boston: Allyn and Bacon; 2003.
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
28. 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]
29. Hale K, Capra S, Bauer J. A framework to assist health professionals in recommending high-quality apps for supporting chronic disease self-management: Illustrative assessment of type 2 diabetes apps. JMIR Mhealth Uhealth 2015;3(3):e87 [FREE Full text] [doi: 10.2196/mhealth.4532] [Medline: 26369346]
30. Huy NP, vanThanh D. Evaluation of mobile app paradigms. 2012 Dec 05 Presented at: 10th International Conference on Advances in Mobile Computing & Multimedia; December 3-5, 2012; Bali, Indonesia p. 25. [doi: 10.1145/2428955.2428968]
31. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
32. Wyatt JC, Thimbleby H, Rastall P, Hoogewerf J, Wooldridge D, Williams J. What makes a good clinical app? Introducing the RCP Health Informatics Unit checklist. Clin Med (Lond) 2015 Dec;15(6):519-521. [doi: 10.7861/clinmedicine.15-6-519] [Medline: 26621937]

Abbreviations

AQEL: app quality evaluation
ICC: intraclass correlations
IRR: interrater reliability
PCA: principal component analysis
MARS: Mobile App Rating Scale
MCAR: missing completely at random

Edited by G Eysenbach; submitted 02.02.17; peer-reviewed by R Eckhoff, B Nezami; comments to author 12.07.17; revised version received 29.08.17; accepted 15.09.17; published 27.10.17.

Please cite as:
DiFilippo KN, Huang W, Chapman-Novakofski KM
A New Tool for Nutrition App Quality Evaluation (AQEL): Development, Validation, and Reliability Testing
JMIR Mhealth Uhealth 2017;5(10):e163
URL: http://mhealth.jmir.org/2017/10/e163/
doi:10.2196/mhealth.7441
PMID:29079554

Original Paper

Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions

Alexander Wilhelm Gorny[1], MBBS, MSc; Seaw Jia Liew[1], MEng; Chuen Seng Tan[1,2], MSc, PhD; Falk Müller-Riemenschneider[1,2,3], MBBS, MSc, MD

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[2]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[3]Institute of Social Medicine, Epidemiology and Health Economics, Charité University Medical Centre, Berlin, Germany

Corresponding Author:
Alexander Wilhelm Gorny, MBBS, MSc
Saw Swee Hock School of Public Health
National University of Singapore
Tahir Foundation Building #10-01
12 Science Drive 2
Singapore, 117549
Singapore
Phone: 65 65164988
Fax: 65 67791489
Email: alexander_gorny@u.nus.edu

Abstract

Background: Many modern smart watches and activity trackers feature an optical sensor that estimates the wearer's heart rate. Recent studies have evaluated the performance of these consumer devices in the laboratory.

Objective: The objective of our study was to examine the accuracy and sensitivity of a common wrist-worn tracker device in measuring heart rates and detecting 1-min bouts of moderate to vigorous physical activity (MVPA) under free-living conditions.

Methods: Ten healthy volunteers were recruited from a large university in Singapore to participate in a limited field test, followed by a month of continuous data collection. During the field test, each participant would wear one Fitbit Charge HR activity tracker and one Polar H6 heart rate monitor. Fitbit measures were accessed at 1-min intervals, while Polar readings were available for 10-s intervals. We derived intraclass correlation coefficients (ICCs) for individual participants comparing heart rate estimates. We applied Centers for Disease Control and Prevention heart rate zone cut-offs to ascertain the sensitivity and specificity of Fitbit in identifying 1-min epochs falling into MVPA heart rate zone.

Results: We collected paired heart rate data for 2509 1-min epochs in 10 individuals under free-living conditions of 3 to 6 hours. The overall ICC comparing 1-min Fitbit measures with average 10-s Polar H6 measures for the same epoch was .83 (95% CI .63-.91). On average, the Fitbit tracker underestimated heart rate measures by −5.96 bpm (standard error, SE=0.18). At the low intensity heart rate zone, the underestimate was smaller at −4.22 bpm (SE=0.15). This underestimate grew to −16.2 bpm (SE=0.74) in the MVPA heart rate zone. Fitbit devices detected 52.9% (192/363) of MVPA heart rate zone epochs correctly. Positive and negative predictive values were 86.1% (192/223) and 92.52% (2115/2286), respectively. During subsequent 1 month of continuous data collection (270 person-days), only 3.9% of 1-min epochs could be categorized as MVPA according to heart rate zones. This measure was affected by decreasing wear time and adherence over the period of follow-up.

Conclusions: Under free-living conditions, Fitbit trackers are affected by significant systematic errors. Improvements in tracker accuracy and sensitivity when measuring MVPA are required before they can be considered for use in the context of exercise prescription to promote better health.

(JMIR Mhealth Uhealth 2017;5(10):e157) doi:10.2196/mhealth.8233

KEYWORDS

heart rate; photoplethysmography; telemedicine; validation studies

Introduction

Sedentary behavior, daily step counts, and moderate to vigorous physical activity (MVPA) have been identified as targets for public health intervention [1]. In response to these findings, the practice of physician-directed exercise prescription emerged as a promising strategy to promote the health benefits of physical activity. Under the tagline *Exercise Is Medicine*, clinicians are advocating nonpharmacological interventions in the management of chronic health conditions such as hypertension and diabetes [2].

The objective of exercise prescription is to assess current levels of activity and guide patients as they increase their levels of exercise. A total of 150 min of MVPA per week is required to sustain health, whereas 300 min is needed to improve health [3]. It is rare for any information on physical activity to be captured in clinical practice, and when patients are assessed for their level of activity, little emphasis is placed on objective measures. In the domain of population research, physical activity questionnaires have for many years been the primary means of measurement [4,5], although more recent studies have used wearable devices to assess activity [6,7].

Today, objective measurement of individual physical activity under free-living conditions has become a reality following the introduction of miniaturized step counters and triaxial accelerometer technology. What started off as ball-in-a-box devices that were clipped to the belt have evolved into sleek wrist-worn gadgets that connect wirelessly to mobile phones and the Internet. Data are captured, logged, analyzed, and displayed within a matter of seconds. Wearable devices, which once had only been used in research settings, are now being marketed directly to consumers on a large scale. A recent systematic review of 22 studies has demonstrated high validity of step count measures among commercial devices [8]. Our own experience has shown remarkable correlation between wrist-worn step counters and scientific devices under free-living conditions [9]. It therefore follows that health care providers and exercise professionals might learn to review and interpret the large amounts of objective physical activity data that patients and clients may have collected incidentally.

Apart from recording daily step counts or total volumes of physical activity, a number of commercial wrist-worn trackers and hybrid watches feature optical sensors that estimate heart rates by means of photoplethysmography. The noninvasive optical probe detects the small variation in light absorption brought about by pulsatile perfusion of tissues [10]. In principle, heart rate measures should offer a number of advantages in activity tracking. First, heart rate monitors outperform accelerometers in capturing non–weight-bearing activities such as cycling and rowing. Second, whereas the latter reflects total volume of activity, heart rate monitors provide information on the relative intensity of activity, allowing MVPA to be more accurately discerned from light activity. Finally, where information on real-time relative intensity is available, there are, in theory, potential applications in the realm of safety monitoring for users at risk of overtraining [11]. The current approach to intensity assessment still relies on tactile carotid or radial pulse rates, a method which is potentially cumbersome or inaccurate in laypersons.

Naturally, there are some important disadvantages using heart rates to approximate physical activity. Heart rate is a vital sign which responds to a multitude of physiological stimuli, including emotional state and illness. Heart rates also tend to exhibit considerable lag in the minutes following the cessation of activity. Moreover, devices that have been shown to track heart rates reliably, such as Polar [12,13] and Actiheart [14,15], by measuring myocardial electric potentials (akin to an electrocardiogram), are cumbersome to wear over extended periods as they need to be strapped across the chest.

We sought to explore whether it would be feasible to adopt a wrist-worn consumer wearable to augment health promotion strategies such as exercise prescription. By monitoring heart rate information through the wearable device, the patient would have a more convenient means to guide the calibration of intensity to attain a specific training target [11]. In turn, these measures could be useful to review compliance during follow-up appointments with the health care provider and make available objective feedback to inform behavior change strategies [16].

The literature offers conflicting information on the utility of wrist-worn heart rate sensors when tested for validity while participating in exercise protocols under laboratory conditions. Authors who chose to cite strong correlation coefficients and low mean percentage errors as validation criteria concluded that wrist-worn devices performed well [17-19]. Others with a stricter definition of accuracy choosing to examine mean bias and levels of agreement [20] concluded that devices performed inadequately. When validated in hospital patients, the devices were found suitable for a subset of patients who were in sinus rhythm [21]. We found only a limited number of validation studies that collected information on minute-by-minute heart rates [22] and daily energy expenditure [23] outside the laboratory. In the time following our data collection efforts, a class action lawsuit was filed against Fitbit Inc, alleging that the devices “consistently mis-record heart rates by a very significant margin, particularly during exercise” [24].

To address some of the existing gaps in the literature, we have conducted this validation study to assess the accuracy of a common wrist-worn heart rate tracking device under free-living conditions and to evaluate the feasibility of including heart rate tracking measures as part of population-based activity monitoring and mobile health interventions.

Methods

Participants

We aimed to recruit 40 members of the university’s staff and students through department-approved internal emails for a pilot study assessing the feasibility of wearable-based observational studies examining physical activity, nutrition, and mental well-being. Out of this pilot study group, a convenience sample of 10 participants would be invited to participate in an additional validation component to examine the accuracy of heart rate measures provided by Fitbit Charge HR (Fitbit, San Francisco CA, USA). Participants could be included if they owned a

compatible mobile phone with a data plan, were aged between 18 and 65 years, and were unlikely to travel abroad over the subsequent 1 month. The following criteria excluded an interested participant from the study: having a severe medical condition that would prevent participation in physical activity, discomfort, or unwillingness to wear multiple devices concurrently and participation in activities or work that would restrict the use of the devices.

Study Procedures and Data Collection

We compiled baseline characteristics for all our participants by means of a self-administered questionnaire. Measures of height and weight were taken using a SECA stadiometer (SECA GmbH, Hamburg, Germany) at our study site.

Each participant was provided a new Fitbit Charge HR (Fitbit) tracker to be worn on the nondominant hand throughout the pilot study. Participants were instructed in the use of the tracker device and the installation of mobile phone apps according to manufacturer's specifications. They were also taught to synchronize the Fitbit tracker periodically.

Each of the 10 participants in the heart rate validation series were also fitted with one Polar H6 heart rate monitor (Polar Electro Oy, Kempele, Finland) worn across the chest. To record the Polar H6 heart rate monitor (Polar) data, these participants were provided with an Actigraph GT3X+ logger (Actigraph) on Bluetooth receiver mode set to sample measures at 10-s intervals and worn on the same wrist as the Fitbit device. The 10 participants were asked to wear all 3 devices for at least 3 and at most 6 continuous hours of nonsleeping activities. They were encouraged to continue pursuing their usual activities, excluding water sports. The Polar and wrist-worn Actigraph devices could be removed before bedtime and returned to the study site over the following days. Participants would continue to use their Fitbit trackers for the remaining 1 month of free-living study. Participants used their personal data plans to run synchronizations with the Fitbit server.

Fitbit heart rate measures were downloaded directly from the Web server using a developer's application programming interface (API) issued by Fitbit. Polar measures were collated from the wrist-worn Actigraph devices. Common wear time for the validation study was defined as every 1-min epoch, which reflected a nonzero heart rate on both devices. For the 1-month period of continuous monitoring, any nonzero heart rate registered by the Fitbit was defined as valid 1-min epoch of wear time. A valid day of wear time was defined as having at least 600 1-min epochs of nonzero counts within 1 calendar day.

Graphical Analysis

Our first dataset comprised one Fitbit heart rate measure for each discreet 1-min epoch, whereas six 10-s measures for the same epoch were available from the Polar device. While the Fitbit API allowed us to review heart rate measures at intervals less than 1 min, the time differences between measures were irregular. Given that Fitbit users would only have access to data logs recorded by minute, we did not attempt to generate our own summary measures for within-minute heart rates. It is worth noting that the literature recommend that photoplethysmographic readings should be averaged over a 60-s duration to obtain a reliable measure [10]. To appreciate the data contributed by individual participants, we generated time series plots of discreet 1-min epochs where Fitbit measures were superimposed onto ranges of Polar 10-s measures. Thereafter, 10-s Polar heart rate measures were averaged for each 1-min epoch. Subsequently, we rank-ordered aggregated discreet epochs by their average Polar measure, divided these epochs by deciles, and constructed box plots to compare average Polar and Fitbit measures. Box plots for the width of within-epoch ranges of 10-s Polar measures were included in this descriptive plot.

Statistical Analysis

Intraclass correlation coefficients (ICCs) were calculated using a mixed effects model assessing for absolute agreement between average 10-s Polar measures, and 1-min Fitbit measures. ICCs were calculated first for overall measures and then for measures stratified by physical activity heart rate zone cut-offs proposed by the Centers for Disease Control and Prevention [25] and by individual participants. Given that the Polar device was chosen as reference, heart rate zones were assigned according to each participant's age and average 10-s Polar measure within discreet epochs. Two Bland-Altman plots were constructed to visually evaluate the overall differences in absolute measures within heart rate zones. A two-by-two table was constructed to estimate the sensitivity, specificity, and positive and negative predictive values for Fitbit devices, correctly identifying MVPA heart rate zones where average 10-s Polar values were considered as reference.

In our second dataset, we compiled all Fitbit heart rate measures obtained in the 1-month free-living study period. The aggregated valid days and epochs of wear time were compiled in a bar chart with superimposed dot and whiskers plots. We tabulated summary statistics for each participant, detailing the number of 1-min epochs spent in respective heart rate zones under free-living conditions.

All statistical analyses were conducted using STATA (Version 13.1, StataCorp LP). Given the exploratory nature of this study, $P<.05$ was chosen as a level of statistical significance. The strength of ICC coefficients was interpreted based on the following definitions: weak ($r<.5$), moderate (.5-.7), and strong ($r>.7$). This study was approved by the institutional review board of the National University of Singapore.

Results

Study Participants

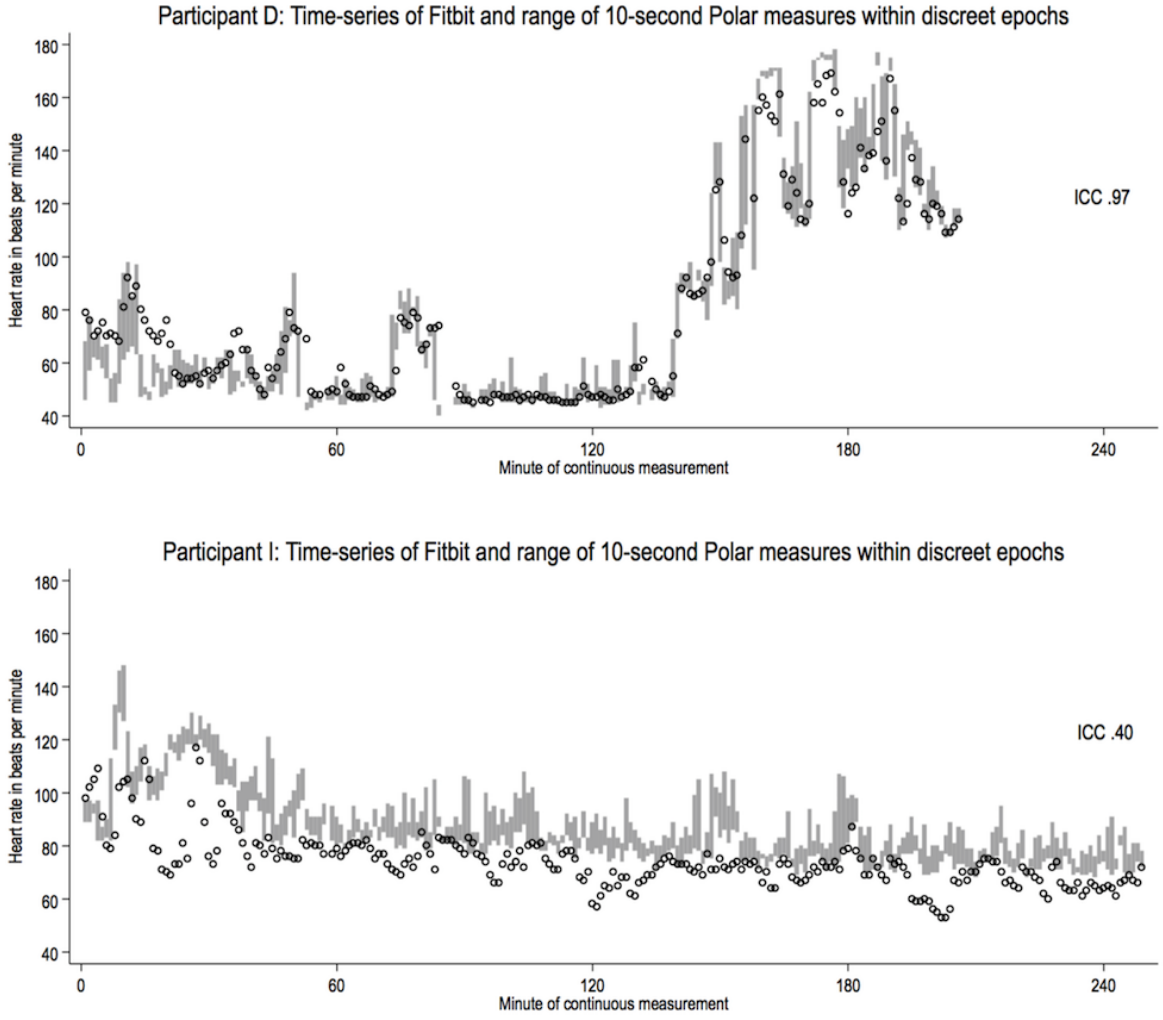
From our pilot study group of 20 males and 20 females, we recruited 3 females and 7 males to participate in the heart rate validation segment. Recruitment and data collection began on November 4, 2015, and the last day of assessment was January 7, 2016. Nine out of 10 participants were students, and their average body mass index was 22.9 kg/m^2 (standard deviation [SD] 3.8). Table 1 describes the characteristics of the final sample of 10 study participants.

Table 1. Sample characteristics (N=10).

Characteristics	Value (N=10)
Gender, n	
Female	3
Male	7
Occupation, n	
Undergraduate	3
Postgraduate	6
Staff	1
Age in years, mean (SD[a])	25.4 (3.7)
Body mass index in kg/m^2, mean (SD)	22.9 (3.8)
Number of valid 1-min epochs contributed, mean (SD)	
Paired measures	250 (95)
1-month follow-up period	38880 (11758)

[a]SD: standard deviation.

Figure 1. Time-series of Fitbit 1-min measures (circles) and Polar 1-min ranges (gray bars) estimating heart rate in beats per minute for participants D and I.



Validation of Fitbit Measures Compared With Polar

Of a total of 2769 possible 1-min epochs, 2509 valid paired readings were identified, with each participant contributing on average 250 (SD 95) epochs. Unpaired readings were treated as missing data and omitted from further analyses. For illustrative purpose, 2 of the 10 time series plots are shown in Figure 1, representing the strongest and weakest measures of intraclass correlation.

The graphical comparison of aggregated epochs in Figure 2 shows how heart rate measures from Fitbit were consistently lower than Polar, whereas the width of 10-s Polar value ranges remained consistent.

Table 2 shows the ICCs and differences between Fitbit and Polar measures. The overall ICC between both devices was strong (.83; 95% CI 0.63-0.91) and ranged from .40 to .97 across participants. The ICC was markedly weaker at MVPA heart rate zones as compared with the low heart rate zone. On average, Fitbit devices measured heart rates that were −5.96 bpm (95% CI −6.33 to −5.60) lower than Polar. Reviewing the differences between participants, we noted that the underestimate was statistically significant in all but 1 participant, who also demonstrated the strongest ICC. Again, the difference between both devices was greater in MVPA heart rate zones. This finding was reproduced in the Bland-Altman plots of measures (Figure 3).

Figure 2. Box-plots providing by-decile comparisons of average Polar (dark gray) and Fitbit (black) measures, width of within-epoch ranges of 10-second Polar measures (white) included (n=2509).

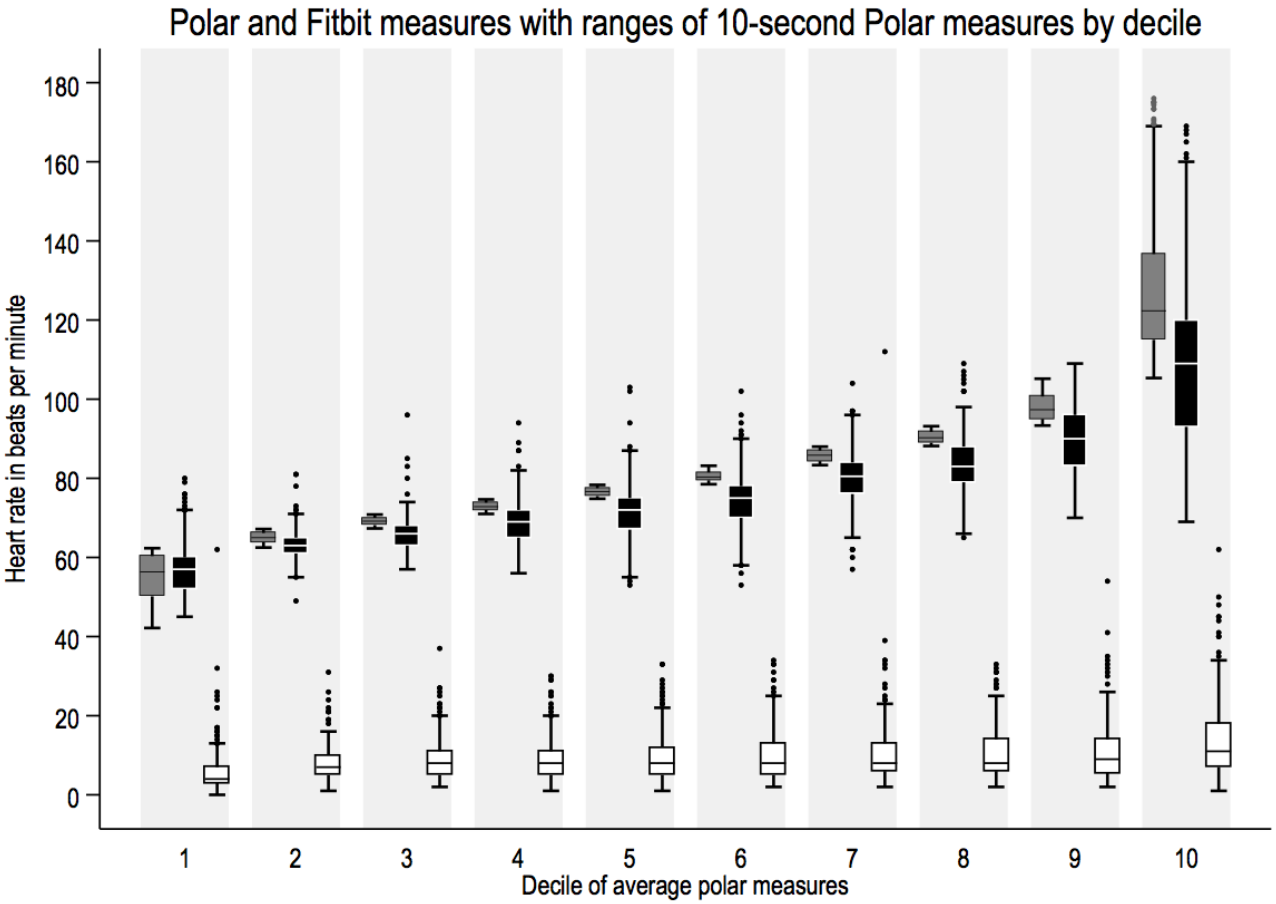


Table 2. Overall and stratified comparisons between Fitbit and Polar of minute-by-minute epochs.

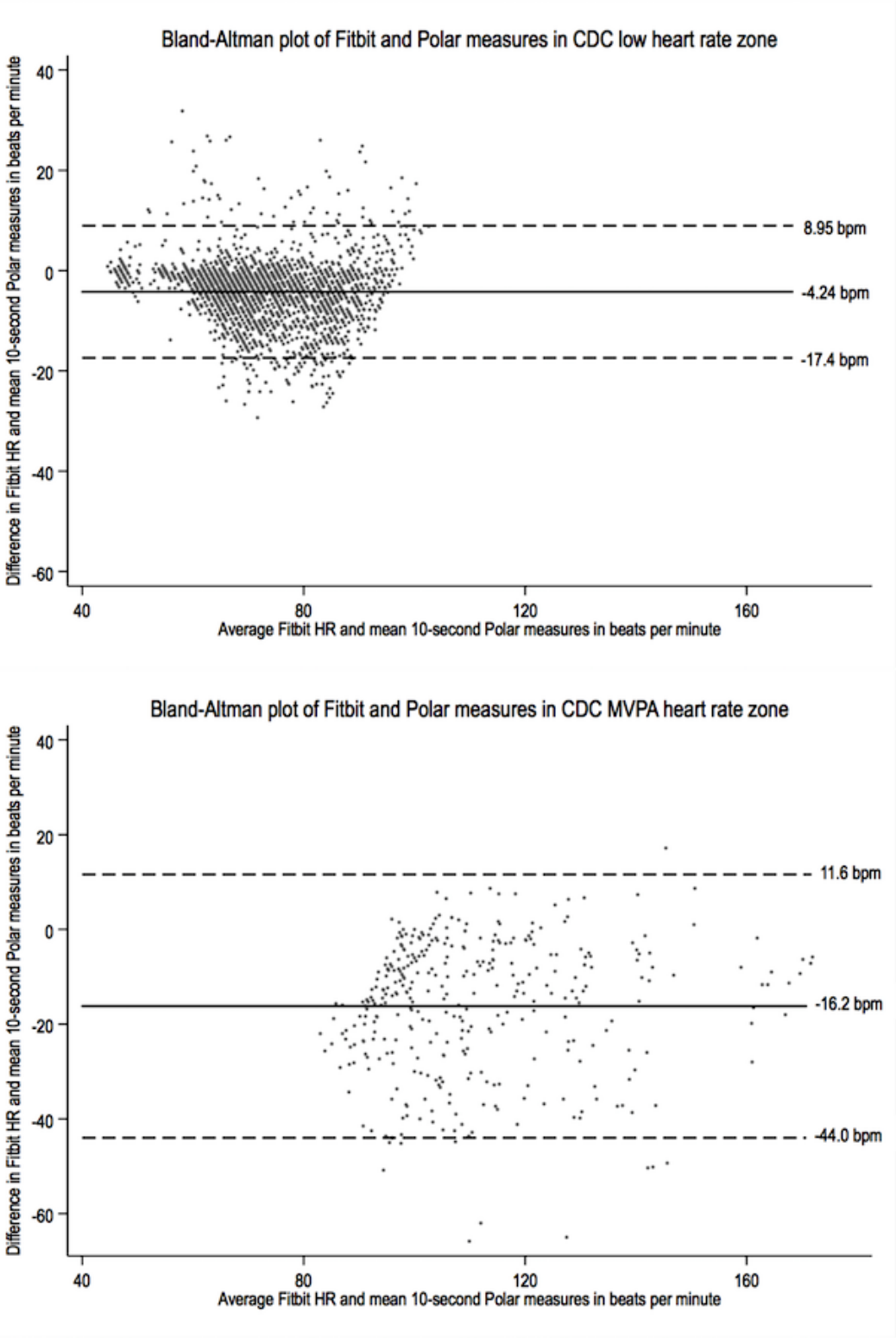
1-min epochs	n	Within-epoch comparison of 1-min Fitbit and average 10-s Polar heart rates (in bpm)		
		Intraclass correlation coefficient[a] (95% CI)	Average difference[b] (95% CI)	*P* value[b]
Overall	2509	.83 (0.63-0.91)	−5.96 (−6.33 to −5.60)	<.001
By Centers for Disease Control and Prevention heart rate zone[c]				
Low (<50%)	2146	.77 (0.55-0.87)	−4.24 (−4.53 to −3.96)	<.001
Moderate to vigorous physical activity (≥50%)	363	.56 (0-0.79)	−16.2 (−17.6 to −14.7)	<.001
By participant				
A	348	.87 (0.69-0.94)	−2.58 (−2.99 to −2.18)	<.001
B	358	.75 (0.50-0.86)	−3.92 (−4.55 to −3.29)	<.001
C	169	.70 (0.28-0.85)	−5.08 (−6.01 to −4.14)	<.001
D	198	.97 (0.96-0.98)	0.25 (−1.08 to 1.58)	.70
E	359	.62 (0.14-0.81)	−6.40 (−7.13 to −5.67)	<.001
F	208	.47 (0-0.74)	−17.4 (−19.5 to −15.4)	<.001
G	320	.72 (0.19-0.88)	−5.05 (−5.62 to −4.47)	<.001
H	228	.81 (0.65-0.88)	−5.20 (−6.45 to −3.94)	<.001
I	248	.40 (0-0.68)	−12.3 (−13.6 to −11.1)	<.001
J	73	.93 (0.88-0.95)	−0.92 (−1.69 to −0.14)	.02

[a]ICCs derived using two-way mixed effects model for absolute agreement.

[b]Statistics derived in paired *t* tests.

[c]Calculated as percent of maximal heart rate (220 bpm—age in years) for discreet 1-min epochs drawing from the average 10-s Polar measure and age of the respective participant.

Figure 3. Bland-Altman plots for paired measures in low (top) and moderate to vigorous physical activity (MVPA; bottom) Centers for Disease Control and Prevention (CDC) heart rate zones.



Within the aggregate 2509 1-min epochs shown in Table 3, only 363 were spent in the MVPA heart rate zone according to Polar measures.

One-Month Continuous Observation

Figure 4 shows that wear time on valid days was consistent, although the number of participants providing valid days of device usage declined over the course of the study period.

On valid days of device usage as shown in Table 4, 24.4% of epochs were classified as nonwear time, 71.7% of epochs fell within the low intensity heart rate zone, and 3.9% of epochs were classified as MVPA. Thus, on average, participants spent 55 min (SD 34) per day in the MVPA heart rate zone.

Table 3. Number of 1-min epochs spent in Centers for Disease Control and Prevention (CDC) physical activity heart rate zone; n=2509 paired observations. Sensitivity and specificity were 52.9% (192/363) and 98.6% (2115/2146), respectively. The positive predictive value for a 1-min epoch categorized as MVPA by Fitbit values was 86.1% (192/223), and the negative predictive value was 92.52% (2115/2286).

Epoch categorization	According to Fitbit values, n (%)		Total, n (%)
	Low	MVPA[a]	
According to average 10-s Polar values			
Low	2115 (98.6)	31 (1.4)	2146 (100.0)
MVPA	171 (47.1)	192 (52.9)	363 (100.0)

[a]MVPA: moderate to vigorous physical activity.

Figure 4. Number of participants with valid days (gray bars) with distribution of aggregated wear time means (boxes) and one standard deviation (whiskers).

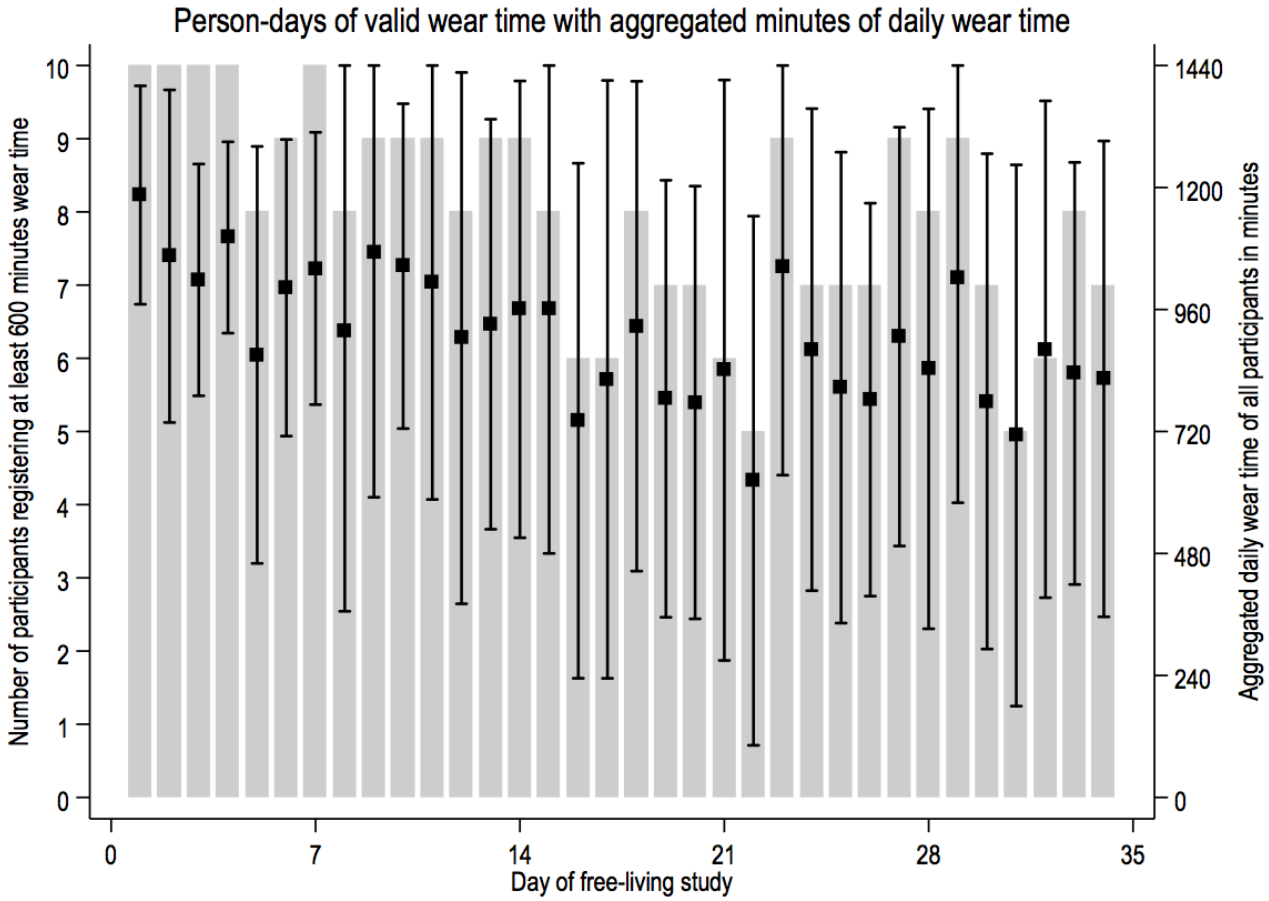


Table 4. Free-living data collection. Over n=270 valid person-days of Fitbit usage.

Device use	Valid days	Total number of epochs	Number of 1-min epochs spent in CDC[a] physical activity heart rate zone, n (%)			Average daily minutes spent in MVPA[b] heart rate zone
			No valid reading	Low	MVPA	
Overall	270	388,800	94,993 (24.4)	278,856 (71.7)	14,951 (3.9)	55 (SD 34)
By participant						
A	21	30,240	13212 (43.7)	16,397 (54.2)	631 (2.09)	30
B	30	43,200	12076 (28.0)	29,729 (68.8)	1395 (3.2)	47
C	26	37,440	13984 (37.4)	20,691 (55.3)	2765 (7.4)	106
D	33	47,520	16136 (34.0)	30,600(64.4)	784 (1.7)	24
E	22	31,680	6754 (21.3)	23,686 (74.8)	1240 (3.9)	56
F	32	46,080	10703 (23.2)	33,593 (72.9)	1784 (3.9)	56
G	34	48,960	2743 (5.6)	44,580 (91.1)	1637 (3.3)	48
H	34	48,960	2694 (5.5)	45,488 (92.9)	778 (1.6)	23
I	30	43,200	11901 (27.6)	27,614 (63.9)	3685 (8.5)	123
J	8	11,520	4790 (41.6)	6478 (56.2)	252 (2.2)	32

[a]CDC: Centers for Disease Control and Prevention.

[b]MVPA: moderate to vigorous physical activity.

Discussion

Principal Findings

In this free-living validation study, we have compiled rich data by relatively convenient means, capturing 2509 1-min epochs of paired data and tens of thousands of unpaired 1-min epochs in the follow-up period. Visual inspection of within-participant heart rate plots showed that there were differences in how well Fitbit readings coincided with Polar ranges. We ascertained an overall strong ICC for absolute agreement between Fitbit and Polar measures that varied markedly between participants and diminished at heart rates that represent moderate to vigorous intensity physical activity. The Fitbit devices identified just over half of MVPA heart rate zone readings correctly.

Our summary measure of a 6 bpm or 7% underestimate of heart rates measured by Fitbit was in keeping with current literature where error estimates have been established under laboratory conditions [18,26,27]. In MVPA heart rate zones, we found an average underestimate of 16 bpm, which was shown to markedly impact Fitbit's ability to correctly identify time spent in MVPA under free-living conditions. This finding of a larger error at higher heart rates is consistent with other studies [26,28]. However, our findings suggest that the systematic underestimation of heart rates might partially be accounted for by differences between participants.

Applying our own measures of sensitivity and specificity to the data obtained during the 1-month follow-up period, we could surmise that close to an hour's worth of MVPA epochs were not captured on any given day. This underestimate of daily MVPA time contrasts with the overestimate observed in a free-living validation study of a wrist-worn tracker of the same brand that measured bouts of activity based on accelerometry [29]. It is also important to note that Fitbit wear time and the number of participants wearing it sufficiently long had decreased considerably even over the 1-month monitoring period. This is consistent with findings from other studies [30] that have also reported considerable drops in compliance with wearable device use over time.

Our free-living validation study into the accuracy of wrist-worn heart rate monitors has several implications on their potential usefulness in monitoring relevant physiological parameters over time and tracking compliance with exercise prescriptions. The results are in keeping with the past studies, which concluded that the device would fare poorly in the calibration of intensity of activity owing to insufficient accuracy. Concerning activity tracking, we found that the devices would fail to recognize one in two MVPA heart rate zone epochs, thus diminishing their value as a means of assessing activity levels objectively. Our follow-up data suggest that device use declined over the course of the study, further complicating potential uses as a compliance monitoring tool.

Overall, our findings have demonstrated that more emphasis should be placed on eliminating systematic error in the tracker measures. Our data showed that errors might be explained in part by putative between-participant differences that would include device fit and skin surface characteristics. Additional mathematical calibration might be appropriate for the trackers to more reliably detect MVPA heart rate zones. As these sensors become increasingly ubiquitous, their potential role in exercise prescription and health promotion merits further evaluation.

Limitations

Our study design was limited to a small sample of mostly male, young adults who did not report significant health issues, thus limiting the generalizability of our findings. The recruitment of female participants was affected by expressed discomfort wearing the Polar H6 chest strap beneath undergarments. In

addition, the measurement period was restricted over a few hours of one day where participants would engage in their normal activities. Although this provided us the necessary information on usual day-to-day life, it resulted in a limited number of paired measures, particularly in the moderate to vigorous heart rate zones. We did not include any form of activity diary in the 1-month period of follow-up, thus limiting our ability to verify the total duration of MVPA accrued. The sensitivity of the photoplethysmographic probe is strongly affected by placement and skin condition [10]. Although we provided advice on proper fit, it is plausible that participants may have foregone scientific accuracy in favor of personal comfort by loosening wrist straps. Due to the logistics of the study, we were unable to ascertain proper fit at the end of the observation period, which could have provided further insights into the apparent interpersonal differences in Fitbit accuracy. Finally, it is important to note that the optical sensors provide a measure of microvascular perfusion, whereas our reference device registers myocardial electric potentials. In practice the arterial pulse rate is often synonymous with the rate of cardiac contractions, but this level of equivalence might be considered inappropriate in the evaluation of photoplethysmographic devices, which are known to be affected by movement and other artifacts that contribute to error rates of up to 8% [10].

Conclusions

The nature of this study was part validation and part exploration. While the overall ICC for absolute agreement appears strong, our data suggest that under free-living conditions, Fitbit Charge HR trackers overall compared poorly against the reference device, especially at higher heart rate zones. Our findings are in line with findings of past studies, which have expressed concern that such devices might not provide adequate information to guide exercise intensity or detect MVPA. Given the nature of our small pilot study with a limited period of observation, further research with a larger sample is warranted to confirm our results.

Acknowledgments

This study was approved by the institutional review board of the National University of Singapore. The project was supported by the Humanities and Social Sciences Seed Fund of the National University of Singapore. The authors would like to thank Yusuf Shamir for his support in managing the data systems and all colleagues and students who participated in the study.

Authors' Contributions

GAW designed the study, recruited the participants, processed the data, performed the data analysis, interpreted the findings, and drafted the manuscript. LJSJ participated in recruitment and data collection. TCS interpreted the findings and reviewed the data analyses. MRF conceived the project, participated in the data analysis and interpretation of findings, edited the manuscript, and supervised the project. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
2. Sallis RE. Exercise is medicine and physicians need to prescribe it!. Br J Sports Med 2009 Jan;43(1):3-4. [doi: 10.1136/bjsm.2008.054825] [Medline: 18971243]
3. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
4. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. Sports Med 2001 Feb;31(2):91-100. [Medline: 11227981]
5. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206; discussion 206 [FREE Full text] [Medline: 12782543]
6. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
7. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]
8. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

9. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017 Feb 24;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
10. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
11. Achten J, Jeukendrup AE. Heart rate monitoring: applications and limitations. Sports Med 2003;33(7):517-538. [Medline: 12762827]
12. Terbizan DJ, Dolezal BA, Albano C. Validity of seven commercially available heart rate monitors. Meas Phys Educ Exerc Sci 2002;6(4):243-247. [doi: 10.1207/S15327841MPEE0604_3]
13. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
14. Barreira TV, Kang M, Caputo JL, Farley RS, Renfrow MS. Validation of the Actiheart monitor for the measurement of physical activity. Int J Exerc Sci 2009:60-71 [FREE Full text]
15. Spierer DK, Hagins M, Rundle A, Pappas E. A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging. Eur J Appl Physiol 2011 Apr;111(4):659-667. [doi: 10.1007/s00421-010-1672-7] [Medline: 20953878]
16. Patel MS, Asch DA, Volpp KG. Wearable devices as facilitators, not drivers, of health behavior change. JAMA 2015 Feb 3;313(5):459-460. [doi: 10.1001/jama.2014.14781] [Medline: 25569175]
17. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016 Apr 25;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
18. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016 May 27;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
19. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2) [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
20. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep 5;15(3):540-547 [FREE Full text] [Medline: 27803634]
21. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: a prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
22. Lee JM, An HS, Kang SK, Kim Y. Digitalcommons.unomaha. 2016. Examining the validity of Fitbit Charge HR TM for measuring heart rate in free-living conditions URL: http://digitalcommons.unomaha.edu/pahppresentations/15/ [accessed 2017-07-23] [WebCite Cache ID 6s9rehpyy]
23. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
24. Plainsite. McLellan et al v. Fitbit, Inc URL: https://www.plainsite.org/dockets/2p5e5b1z6/california-northern-district-court/mclellan-et-al-v-fitbit-inc/ [accessed 2017-07-22] [WebCite Cache ID 6s9pr4cIE]
25. CDC. Target heart rate and estimated maximum heart rate URL: https://www.cdc.gov/physicalactivity/basics/measuring/heartrate.htm [accessed 2017-06-14] [WebCite Cache ID 6rDxXmPLq]
26. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the future patient trial, phase I. Sensors (Basel) 2017 Jan 22;17(1) [FREE Full text] [doi: 10.3390/s17010211] [Medline: 28117736]
27. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
28. Jo E, Dolezal BA. Lieffcabraser. Validation of the Fitbit surge and charge HR fitness trackers URL: http://www.lieffcabraser.com/pdf/Fitbit_Validation_Study.pdf [accessed 2017-06-14] [WebCite Cache ID 6rDy2pG2m]
29. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
30. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]

Abbreviations

API: application programming interface
CDC: Centers for Disease Control and Prevention
ICC: intraclass correlation coefficient
MVPA: moderate to vigorous physical activity
SD: standard deviation
SE: standard error

Edited by G Eysenbach; submitted 22.06.17; peer-reviewed by M Lang, D Maslove; comments to author 15.07.17; revised version received 08.08.17; accepted 06.09.17; published 20.10.17.

Please cite as:
Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F
Fitbit Charge HR Wireless Heart Rate Monitor: Validation Study Conducted Under Free-Living Conditions
JMIR Mhealth Uhealth 2017;5(10):e157
URL: http://mhealth.jmir.org/2017/10/e157/
doi:10.2196/mhealth.8233
PMID:29055881

Original Paper

Well-Being Tracking via Smartphone-Measured Activity and Sleep: Cohort Study

Orianna DeMasi[1,2], BS; Sidney Feygin[3], MS; Aluma Dembo[4], PhD; Adrian Aguilera[5,6], PhD; Benjamin Recht[1], PhD

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, United States

[2]Berkeley Institute of Data Science, University of California, Berkeley, Berkeley, CA, United States

[3]Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA, United States

[4]Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA, United States

[5]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[6]Zuckerberg San Francisco General Hospital, Department of Psychiatry, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Orianna DeMasi, BS
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
593-5 Soda Hall
MC-1776
Berkeley, CA, 94720
United States
Phone: 1 5107769028
Email: odemasi@eecs.berkeley.edu

Abstract

Background: Automatically tracking mental well-being could facilitate personalization of treatments for mood disorders such as depression and bipolar disorder. Smartphones present a novel and ubiquitous opportunity to track individuals' behavior and may be useful for inferring and automatically monitoring mental well-being.

Objective: The aim of this study was to assess the extent to which activity and sleep tracking with a smartphone can be used for monitoring individuals' mental well-being.

Methods: A cohort of 106 individuals was recruited to install an app on their smartphone that would track their well-being with daily surveys and track their behavior with activity inferences from their phone's accelerometer data. Of the participants recruited, 53 had sufficient data to infer activity and sleep measures. For this subset of individuals, we related measures of activity and sleep to the individuals' well-being and used these measures to predict their well-being.

Results: We found that smartphone-measured approximations for daily physical activity were positively correlated with both mood (P=.004) and perceived energy level (P<.001). Sleep duration was positively correlated with mood (P=.02) but not energy. Our measure for sleep disturbance was not found to be significantly related to either mood or energy, which could imply too much noise in the measurement. Models predicting the well-being measures from the activity and sleep measures were found to be significantly better than naive baselines (P<.01), despite modest overall improvements.

Conclusions: Measures of activity and sleep inferred from smartphone activity were strongly related to and somewhat predictive of participants' well-being. Whereas the improvement over naive models was modest, it reaffirms the importance of considering physical activity and sleep for predicting mood and for making automatic mood monitoring a reality.

(JMIR Mhealth Uhealth 2017;5(10):e137) doi:10.2196/mhealth.7820

KEYWORDS

depression; mobile health; smartphones

Introduction

A goal of personalized medicine is to tailor treatments to individuals based on their needs. To aid the tailoring of treatments, it is necessary to monitor an individual's state of well-being and to evaluate whether they are responding to a treatment [1,2]. However, monitoring can be a tedious and expensive process and, as a result, can yield low adherence [3].

To overcome low patient adherence, automatic monitoring can be employed in the treatment of mental health disorders, such as depression and bipolar disorder, which benefit from monitoring symptoms over time to identify symptom relapse and to possibly prevent symptoms because of higher self-awareness [4].

The proliferation of personal electronics has enabled continuous personal monitoring [5]. For example, activity recognition has enabled tracking to monitor physical exertion and sleep patterns [6]. Recent studies have started examining whether these smartphone-measured behavioral patterns can be used to infer and then automatically track signals that are not explicitly measured by the smartphone, such as mental well-being.

Many studies have looked at inferring measures of mental well-being from smartphone-measured behavioral patterns [7]. In particular, researchers have considered using measures of location and mobility from global positioning system (GPS) logs to infer depression [8-10], bipolar state [11], stress [12], and well-being measures related to schizophrenia [13]. These studies have shown that daily self-reported levels of stress are related to geospatial activity and sleep [12] and that mobility data can improve predictions of whether a participant is happier or less depressed than usual [8,10] and their bipolar state or transition between states [11]. Researchers have also found that regularity of an individual's daily mobility is significant when predicting depression symptom severity [9,14].

Additional studies have explored the relationships of social signals such as phone usage, call logs, and SMS (short message service) logs with well-being. Two recent studies found that phone usage measures were correlated with depressive symptom severity [9,14]. Another study found that using social signals such as emails, SMS and call logs, Internet usage, app usage, and location frequency was predictive of mood and energy when previous observations of mood and energy were included [15]. However, a similar follow-up study was unable to reproduce these results. This follow-up study did not find sophisticated models considering high accelerometer activity, call and SMS logs, screen events, app usage, and number of images taken to be better than guessing each individual's well-being [16].

Whereas this body of literature has established that relationships between measures of mental well-being and smartphone-measured behaviors may exist, the above literature has not focused extensively on physical activity in uncontrolled environments (ie, outside a lab without constraints on participants, such as where the phone must be located). For example, studies have explored predicting bipolar states and state transitions via accelerometers on small populations [11] or mood in constrained environments where the phone had to be in a fixed position [17,18] or activities had to be performed in a lab [19]. One study looked at a measure of total daily physical activity and sleep (as measured with multiple sensors) but within the context of stress and not well-being more broadly, and it did not attempt to predict well-being [12].

Despite these few studies' limited focus on activity and sleep, there is a body of literature external to mobile health (mHealth) that has established a strong relationship of better mood with increased activity [20-24] and sleep quality [25,26]. There is also mounting evidence that a smartphone accelerometer measures physical activity to a sufficient extent to be useful for monitoring well-being. Several studies have demonstrated that individuals' sleep and physical activity can be somewhat accurately tracked with smartphones [27] and activity recognition [28-30], respectively. As a result, it seems probable that an individual's activity and sleep, as tracked by their smartphone's accelerometer, could be related to and potentially predictive of their mood and well-being more broadly.

If possible, tracking mental well-being with an accelerometer could have benefits over using other sensors. For example, an accelerometer could provide more privacy than previously considered sensors, such as GPS location [8-12] and call logs [13,15,16]. Another advantage to using an accelerometer is that the sensor is always available when the phone is turned on, including when the individual's phone is out of service or, for example, in a tunnel. Whereas accelerometers embedded in a wearable device might have more potential to accurately track activity, smartphones are more ubiquitous and thus more realistic for long-term tracking.

Here, we are interested in focusing on and better understanding the relationships of physical activity and sleep, as measured by a smartphone accelerometer, with emotion for improving automatic mood tracking. We are particularly interested in understanding whether the relationships are predictive, especially from data collected with ordinary participant-owned smartphones in unconstrained environments (ie, not imposing constraints on participants about where they need to keep the phone or whether they need to have a special device with an accelerometer attached to their body). To explore these research questions, we conducted a field study, extracted measures of physical activity and sleep from smartphone accelerometer logs, related these measures to participants' self-reported well-being, and attempted to infer participants' well-being with classification and regression models. We expect that increased physical activity and better sleep quality will be related to improved self-reported mood and well-being.

Methods

Field Study

We recruited 106 participants from the university community through the Experimental Social Science Laboratory (XLab) for an 8-week field study to pilot methods. Participants were eligible if they owned an Android smartphone, were native English speakers, were undergraduate students, and agreed to the consent form. The study was approved by the University of California, Berkeley Internal Review Board. The participants were asked to take an entry survey, respond to daily well-being prompts on their smartphone, allow passive collection of sensor data from their smartphone, and take an exit survey.

Data Collection

Data were collected from participants through a custom Android app that used the Funf Open Sensing Framework [31]. This app was installed by participants before the study period and collected both passive sensor data as well as daily participant

input. The participants were instructed and reminded to uninstall the app at study completion.

To quantify well-being, we followed prior studies and asked participants to repeatedly fill out a 2-question survey on their phone. Participants could enter information about their state on two 9-point Likert scales—one for energy and one for mood. Scales were labeled with opposite poles, such as unhappy to happy and unenergetic to energetic. Participants could select the specific words from short lists of relative synonyms for each pole, such as unhappy, negative, sad, bad versus happy, positive, good. Participants were queried for their state 4 times a day. Each of the four daily surveys occurred at a random time within a predefined period between 8 AM and 10 PM. The purpose of randomizing within periods was to ensure distribution of surveys throughout the day without having participants anticipate them. All responses given in a day were averaged into a daily level of perceived mood and energy.

To measure activity, we sampled the smartphone's accelerometer for intervals of 3 seconds every 5 minutes. These data were collected continuously from the time the app was installed. There were compatibility issues with phone models and network connections, hence, the amount of data collected on each subject varied. Quality of accelerometers also varied between phone models, which contributed to variance in the amount and quality of data collected on each individual. Some of the difficulties we encountered with sensor data collection included entirely missing observations, nonuniform readings during an observation interval, and insufficient duration of sampling, that is, less than 3 seconds. Participants were excluded from the analyses if they did not have complete data (well-being responses and activity readings) for at least 14 days of the study.

Data Processing

Preprocessing

The smartphones' 3-axis accelerometers measured the acceleration of the device in three directions. Following prior work, we considered the magnitude of the acceleration minus gravity [32]. Gravity for each segment was estimated as the average of coordinates in each of the directions. To account for irregular sampling and to reduce noise in the sensor readings during a sampling interval, we interpolated the available data points and took regular sampling from the interpolation. Quadratic and cubic splines gave irregularities with missing readings; thus, a linear spline was identified as performing the best. This regular sampling allowed us to compute discrete Fourier transforms on the approximated signal and approximate the spectral density using Welch's method, that is, averaging between Fourier transforms on multiple overlapping segments of the full observation window.

Activity Inference

We inferred activity from features summarizing the orientation-invariant magnitude of acceleration deviation and the spectral density of the magnitude of deviation of acceleration. The acceleration deviation was computed by subtracting the estimated gravity from all readings in the interval. This approach was taken to allow for more fine-grained analysis of movement than is presented here. Much prior work with accelerometers, predicting both mental well-being [11,17,19] and activity [28-30], utilized features on coordinate-wise acceleration. However, such approaches were not applicable here, as our participants' phones were not in a fixed position during the study. We followed prior work that considered features on the magnitude and power spectrum of the magnitude of acceleration during the sample period [30]. The features we used were the average and standard deviation of the magnitude of acceleration and the dominant frequency, entropy of the normalized power spectrum, power in the high frequencies, medium frequencies, and low frequencies of the power spectrum of the magnitude of the acceleration. These eight features were used to fit two logistic regression classifiers. One classifier was trained to identify when the phone is *still* or set down; the second classifier identified *activity* such as walking, running, or pedaling a bicycle. We did not use a classifier to explicitly identify the phone being in a vehicle, such as a car, bus, or train. We did not find a classifier to be reliable enough, given the many states a vehicle can assume, for example, idling, accelerating, and traversing a smooth or bumpy road. Such a task was also of uncertain necessity because participants do not necessarily exert extra energy while riding in transportation and thus vehicle activity was less likely to correspond to elevated mood from physical exertion.. As a result, we focused this study on measures of physical activity and sleep. The goal of these two classifiers was to quantify how long the phone was set down at night, and the subject presumably sleeping, and how long the participant was physically active during the day. These classifiers were trained on an auxiliary activity-labeled dataset that was collected with the same smartphone app and data processing pipeline. The classifiers achieved 80% to 95% accuracy on held out subjects from the training dataset.

Measure Extraction

Sleep Duration

Sleep duration was estimated as the length of the longest period during which the participant was not physically active, starting after 9 PM the prior evening. This period was calculated by looking at the longest contiguous series of observations when the accelerometer data predicted that the participant was *not active* and taking the duration of that period. Whereas this approach likely overestimates the duration of sleep, it should be representative of a period of passivity or evening rest and is preferable to the highly noisy alternative of considering the duration for which the phone was predicted to be still during the evening.

Nighttime Stillness

Sleep disturbance, or nighttime stillness, sought to capture sleep disturbance during the time when each participant's phone was most likely to be set down and the participant presumably asleep, based on their typical behavior. This measure was considered to be the fraction of time that a participant was still during their median period of late evening or when their phone would typically be still, based on their behavior during the study. The period of late evening was defined for each participant by first considering the longest contiguous set of observations during which the phone was predicted to be set down, starting after 9

PM for each day of the study. The median time that this period started, or presumably the phone was set down, for each day of the study defined the beginning of period, and the median time that the contiguous *still* observations ended on each day of the study was considered the end of the period of late evening.

Table 1. Daily measures of activity and sleep and how they were calculated.

Type of measure	Measure	How it was measured and calculated
Time	Day of study (semester)	Coded as the number of days since the first day of the study.
	Day of week	Ordinal variable coded Monday (0) through Sunday (6).
Sleep	Sleep duration	Longest contiguous time that the participant was not physically active starting after 9 PM.
Activity	Daytime activity	Fraction of time a participant was physically active during the median active period. The median active period is the time between the median hour the participant became physically active during each day of the study and the median hour that the participant stopped being active during the study.
	Nighttime stillness	Fraction of time the phone was predicted to be still, that is, set down, during the median still period. The median still period was calculated over the course of the study to be the median hour that the longest contiguous still period started and the median hour it stopped.

The nighttime stillness measure for each day of the study was the fraction of observations on that day of the study, which occurred during the late evening period and was predicted to be *still*.

Daytime Activity

For a measure of daily physical activity, we consider the daytime activity, which was the fraction of time that a participant was predicted to be physically active during their active period or the period of the day that we would expect each participant to be active, given their typical behavior during the study. The active period of the day was determined by first looking at the longest contiguous set of observations when the phone's predicted behavior was *not-physically active*, starting after 9 PM. The median time across all the days of the study when this physically not-active period began was considered as the end of the active period, and the median end time of the not-active period was considered the beginning of the participant's typical active period. The *daytime activity* measure for each day of the study was then the fraction of time that the participant's phone predicted (with the models discussed previously) that the participant was *physically active* during the participant's active period.

Day of Study

Following prior work, we coded the day of the study as the number of days that had elapsed since the first day of the study [12]. This measure is important to account for potential participant fatigue, and also to represent the progression of the academic semester, which may have had an effect on the participants.

Weekday

The day of the week, and thus the potential effect of weekends, was accounted for by coding weekdays with an ordinal variable from 0 to 6, Monday through Sunday (Table 1).

Analyses

Relating Measures to Well-Being

The first set of analyses sought to study the relationship of activity, sleep, and time on daily well-being. To account for the repeated measures design and missing data, we used mixed-effects linear models to relate reported average daily well-being measures to daily behavior measures [33]. We started with a maximal random-effects structure for each well-being measure to allow for individual variation and increase generalizability. Due to lack of initial convergence of the model, we followed suggestions in prior work to look at the covariance of the partially converged model and remove the variable in minimum variance from the random-effects structure [34]. Using this procedure, we removed the measure of sleep disturbance, *nighttime stillness*, from the random-effects structure when modeling mood and removed the scaled ordinal variable coding the day of the week when modeling energy. After this step, both models converged. Activity and sleep measures were centered and normalized within individuals, and time measures were scaled between 0 and 1 before fitting the models to compare the relative sizes of effects.

To ensure the value of the model with maximally justified random-effects structure, we fit two additional models: (1) a model with only random intercepts and no additional random-effects or fixed-effects and (2) a model with fixed-effects and a random intercept only. Model fit was assessed with chi-square tests on the log likelihood values of different models. Model assumptions were visually checked. The linear mixed-effects models and analyses were carried out in the R programming language and environment [35] using the lme4 [36] and lmerTest [37] software packages.

Predicting Well-Being

The second set of analyses assessed whether the relationships between daily mood and the activity, behavior, and time features were strong enough to be predictive. To do this, we attempted two tasks. The first task was to predict whether a participant was having a bad day, that is, whether their well-being was lower than their median-reported well-being. Only participants with sufficient observations of each class (at least 5 fine days and 5 bad days) were included in the analysis. The second task was to predict a participant's level of well-being.

Prediction Models

For the first task, predicting whether a participant was having a worse-than-usual day, we used logistic regressions with an L1 and an L2 norm penalty as well as support vector machines

(SVMs) and random forests [38,39]. For predicting the daily level of well-being, we used a linear regression model with the elastic net penalty [40] in addition to an Epsilon-Support Vector Regression and random forests. These models were used on individuals' data to build *personal models*, rather than pooling all individuals' data into a *global model*. Personal models were used because they have been shown to be the most successful approach to predicting individuals' responses [13]. Mixed-effects models help to model behavior within the population as a whole while taking into account the fact that individuals have different behavior, but personal linear models are a best-case scenario for predicting individuals' behavior from their own data.

Prediction Framework

For both prediction tasks, we evaluated prediction accuracy with leave-one-out cross-validation on personalized models, that is, we trained a model on all but one of a participant's data points, evaluated the model accuracy on the held-out observation, and then averaged accuracy across observations. The penalty weights hyperparameters were set with leave-one-out cross-validation on the training data and scanning a variety of penalty weights. The predictive analysis was performed in Python with the scikit-learn library [41].

Model Evaluation: User Lift

The accuracy of predicting whether an individual was having a good day was quantified by prediction error or the percentage of observations that were incorrectly predicted. The accuracy of predicting the level of well-being on a given day was quantified by root-mean-square error, which is the square root of the average squared distance of a prediction from the true value. We report the accuracy of predictions compared with the accuracy of predicting each participant to be at their most common state. This measure is called *user lift*; it is the increase in accuracy, or decrease in error, that the model has relative to always predicting an individual to be at their most common state [42]. By comparing a model with each participant's baseline, user lift reveals how much better a model is doing than guessing a participant to always be at their usual state. We then used permutation tests to assess whether user lift was significantly positive across the participants, that is, whether the models were significantly better than always guessing a participant to be at their most common state, as permutation tests are reported to be more reliable than paired nonparametric tests [43,44].

Results

Participation

Of the 106 participants recruited, 87 installed our app; 57 completed the study, that is, completed the exit survey at the end of the 8-week study period. However, there were only sufficient data on 53 participants to include in the analyses. Baseline characteristics of individuals included and excluded from the analyses are shown in Table 2 and indicate that similar populations were included and excluded from the analyses. Whereas some attrition was because of participation waning over the 8-week study period, there was also attrition as a result of technical difficulties and app compatibility issues on older phones.

Table 2. Participant baseline characteristics. Averages across individuals are reported with standard deviations in parenthesis, except where indicated. Where appropriate, numbers represent the average across individuals of averages within individuals.

Participant measure	Included participants with exit survey (n=47)	Included participants with no exit survey (n=6)	Excluded participants because of insufficient data (n=53)
Age[a]	19.83 (1.99)	20.33 (1.60)	20.80 (4.13)
Female (number)[a]	26	3	28
BDI-20[b] score (entry)[a]	11.14 (9.27)	7.33 (3.54)	12.61 (7.20)
BDI-20[b] score (exit)[a]	11.98 (12.00)	N/A	N/A
Median mood rating	5.17 (1.63)	5.83 (0.90)	5.44 (1.44)
Median energy rating	5.60 (1.27)	6.67 (0.94)	5.98 (0.80)
Number of emotion surveys completed	160.51 (44.42)	139.33 (55.01)	30.25 (50.97)
Number of days with emotion ratings	49.45 (8.27)	44.00 (11.06)	10.49 (15.99)
Reported typical sleep duration in hours (from exit survey)[a]	6.88 (1.35)	N/A	N/A
Average duration of inactive period in hours (sensed *sleep duration*)	8.79 (1.22)	8.56 (0.48)	N/A
Number of times per month a participant exercised (from exit survey)[a]	4.24 (5.04)	N/A	N/A
Average minutes active per day (sensed *daytime activity*)	118.78 (32.67)	151.25 (59.68)	N/A
Number of days with sensed activity and mood input	38.60 (9.15)	40.00 (9.64)	3.36 (5.15)

[a]Indicates measures averaged only over submitted responses, as entry and exit survey questions were optional.

[b]BDI-20 indicates optional self-reports to 20 questions of the Beck's Depression Inventory (the question related to suicidal ideation was omitted).

Table 3. Results of fixed-effects for linear mixed-effects model of mood level from smartphone-measured and time variables. The measure for nighttime stillness was excluded from the otherwise maximal random-effects structure.

Fixed-effect	Estimate	Standard error	*t* value (degrees of freedom)	*P* value
Mean mood (intercept)	5.056	0.174	28.973 (49.0)	<.001
Day of study (semester)	−0.059	0.261	−0.226 (47.0)	.82
Day of week (coded 0-6, Monday-Sunday)	0.040	0.076	0.528 (257.0)	.60
Sleep duration	0.072	0.030	2.451 (52.0)	.02
Daytime activity	0.097	0.032	3.062 (50.4)	.004
Nighttime stillness	0.040	0.026	1.528 (1881.5)	.13

Table 4. Checking model fits for linear mixed-effects model of mood.

Model name	Akaike information criterion	Bayesian information criterion	Log likelihood	Chi-square value (degrees of freedom)	*P* value
Random intercept only	6522.0	6538.8	−3258.0		
Fixed-effects with random intercept only	6508.8	6553.7	−3246.4	23.2 (5)	<.001
Maximal random-effects structure	6322.0	6445.4	−3139.0	214.8 (14)	<.001

Relationship of Sensor Data With Well-Being

From linear mixed-effects models, we found significant positive relationships of daytime activity and sleep duration with daily mood; when participants get more sleep and more daily activity they tend to report better moods (Table 3). Daytime activity has a stronger relative effect than sleep duration. Of note is that nighttime stillness (sleep disturbance) is not significant. This lack of significance could imply that the measurement is too noisy and that more work is needed to reliably measure sleep disturbance with a smartphone. The model with the maximal random-effects structure better accounted for the variance across individual participants than the random intercept only model (Table 4). The main effects also remained significant, even when accounting for individual differences.

We also found a significant positive relationship of daytime activity with daily perceived energy level (Table 5). The relation for sleep, though negative, is not significant, revealing a potentially different relationship between the two emotions (mood and energy) with sleep.

Day of the week has a significant positive fixed-effect but had to be removed from the random-effects structure following prior suggestions about how to handle lack of model convergence [33].

Table 5. Fixed-effects for a mixed-effects linear model relating daily energy level from smartphone-measured and time variables. The ordinal variable for weekday was excluded from the near-maximal random-effects structure.

Fixed-effect	Estimate	Standard error	*t* value (degrees of freedom)	*P* value
Mean energy (intercept)	5.686	0.184	30.857 (53.9)	<.001
Day of study (semester)	−0.304	0.233	−1.303 (49.4)	.20
Day of week (coded 0-6, Monday-Sunday)	0.196	0.067	2.912 (1876.2)	.004
Sleep duration	−0.027	0.031	−0.858 (57.7)	.39
Daytime activity	0.182	0.039	4.673 (49.6)	<.001
Nighttime stillness	0.024	0.030	0.810 (50.4)	.42

Table 6. Checking model fits for linear mixed-effects model of energy.

Model name	Akaike information criterion	Bayesian information criterion	Log likelihood	Chi-square value (degrees of freedom)	*P* value
Random intercept only	6284.2	6301.0	−3139.1		
Fixed-effects with random intercept only	6196.1	6240.9	−3090.0	98.1 (5)	<.001
Maximal random-effects structure	5972.5	6095.9	−2964.2	251.6 (14)	<.001

This effect for day of the week indicated that participants collectively felt more energy at the end of the week, and there is not sufficient evidence to support the idea that weekday affected participants differently.When we changed the variable encoding weekday to a binary variable indicating a fixed weekend of Saturday and Sunday versus the rest of the week, as has been suggested in related work [14], this relationship did not remain significant. An interaction term between a weekend indicator and daily activity was similarly not found to be significant. This lack of significance as a binary variable could be a result of weekends being less defined in our undergraduate population, some of whom may or may not have classes on Friday and thus have had extended *weekends*. The lack of significance could alternatively result from insufficient observations of weekends for each participant. Again, sleep disturbance is not significant, further indicating that there might be too much noise in the variable measuring sleep quality. The model with the maximally justified random-effects structure accounted for significantly more variation across participants than having only a random intercept (Table 6).

Predicting Well-Being From Sensor Data

The activity, sleep, and time measures described above were also used to predict daily well-being scores. Whereas mixed-effects models were used to understand relationships of activity and sleep measures with well-being within the population, personal models (linear and nonlinear) were used as a maximally personalized and thus somewhat best-case approach for predicting individuals' well-being [13]. The *user lift*, or improvement of model predictions over a baseline is reported (Table 7). The user lift is the increase in accuracy (or decrease in error) that a model has relative to always predicting an individual to be at their most common state. User lift compares a model's accuracy with a participant's baseline; thus, it quantifies how much better a model is performing than the most reasonable constant prediction for each participant.

In general, it was difficult to predict individuals' well-being on a daily basis with the given information only being about their activity and sleep (Table 7). On average, the best models were able to improve prediction of good and bad mood and energy by 5.44% and 4.92%, respectively. The model prediction performance presented in Table 7 is for linear models (penalized logistic regression and elastic-net penalized linear regression), as those models were found to return higher accuracy than the nonlinear SVMs and random forests. Whereas there was considerable variation in predictability across individuals, permutation tests reveal that user lift was significantly greater than 0, that is, the models were better than naively always predicting each participant to always be at their most common state.

Table 7. Statistics on linear models predicting daily well-being from activity measures. Whereas the models provide an improvement overall, there is a range in the ability to model individuals. The *P* values are for permutation tests, checking whether user lift is greater than 0, that is, whether models are significantly more accurate than always predicting each individual to be at their most frequent state.

Problem (model)	Well-being measure	Average user lift	Minimum user lift	Maximum user lift	*P* value
Good or bad day (penalized logistic regression)	Mood (Prediction error)	5.44%	−21.74%	35.00%	.001
	Energy (Prediction error)	4.92%	−22.73%	39.39%	.008
Daily average (linear regression with elastic net)	Mood (RMSE[a])	0.026	−0.232	0.48	.08
	Energy (RMSE)	0.048	−0.169	0.575	.01

[a]RMSE: root-mean-square error.

Discussion

Principal Findings

We found that increased daily activity, as tracked with a smartphone's accelerometer, positively correlated with participant-reported mental well-being over time. Whereas a positive correlation of activity and well-being has been substantiated in literature external to mHealth [20-24], we have shown that smartphones measure individuals' daily activity to a sufficient level of accuracy to measure this relationship in everyday life. Although the potential for this result has been shown in environments where constraints were placed on the participants [11,17-19], we found this relationship present when no constraints were placed on participants. Previous work did not find a significant correlation of the total activity in a 24-hour day with stress [12], which could indicate the need for distinguishing daytime activity from nighttime activity, as we have done, or indicate that physical behavior has unique effects on different emotions, which we have observed by considering mood and energy separately.

We also found that a simple measure of sleep duration derived solely from accelerometer data was significantly positively correlated with mood. However, it was not significantly correlated with perceived energy, which supports the idea that there are different relationships between different emotions and physical behaviors. We did not find a significant correlation of either mood or energy with our measure of smartphone-measured sleep disturbance. This may imply that the measure did not sufficiently describe sleep quality and that more work is needed to monitor sleep quality in a sustainable manner. It is possible that a more sophisticated method for predicting sleep, such as the method found in prior works, would allow for a finer measure of sleep disturbance [27].

When we used the activity, sleep, and time measures to predict individuals' well-being, we found modest but significant improvement over naive baseline models. It is important to emphasize that there was a range in our ability to predict

individuals' well-being from their activity and sleep behavior. This range highlights the need for tracking approaches that tailor to the user. However, it is unclear whether this effect is the result of a range in how thoughtfully individuals responded with their state, phone usage, data quality and quantity, or the strength of well-being and activity relationship between individuals.

Limitations

A limitation of this study is that participants' self-reported well-being is subjective, and the population was not clinically assessed. However, the measures of well-being that we used have been widely used and prior research has found simple single-scale measures to be related to longer clinical assessments [45]. Whereas a better measure of well-being could be a longer survey, such a measure would incur significant participant fatigue and likely decrease the duration of participation.

Whether all of the participants' relevant activity was tracked with smartphones during the study is another concern. There are limitations to activity recognition, especially when the smartphone is not in a fixed position, a participant is performing a nonstandard activity, or the phone is set down, for example, left in a gym locker. However, the study cohort retrospectively reported little vigorous exercise during the study period (Table 2); thus, the underestimation of vigorous exercise is likely to be minor. Such limitations could possibly be partially mitigated with location tracking, but time at a location is not necessarily representative of activity, and poor GPS sensitivity would remain a challenge. Wearables may provide a better facsimile of an individual's behavior when they are worn, but they have notorious compliance limitations that smartphones do not suffer.

Another limitation was the sample size and lack of clinical population. Some of the individuals in our study cohort did report elevated levels of depressive symptoms in the entry and exit survey. However, the cohort is not necessarily representative of a population with clinically diagnosed mood disorders. Depressed individuals often are less active than the general population, but even small increases in physical activity can improve symptoms [46].

Conclusions

This study examined the extent to which smartphones' accelerometers can contribute to passively tracking individuals' mental well-being in everyday life. We have found that smartphones measure activity and sleep with sufficient accuracy to reproduce prior findings of significant relationships between activity and sleep with mood. Whereas models have a modest, though significant, improvement over naive baseline models in general, the range in predictive capability implies that more work is needed to tailor mood- and depression-tracking apps to individuals.

Our results support the promise for smartphones to be used in sophisticated and long-term monitoring of patients' well-being. Because smartphone use is high and their presence ubiquitous, the ability to use a smartphone for tracking mental well-being could have a huge impact on mental health care. Smartphone monitoring may improve self-management via smartphone apps, thereby making care more affordable and thus accessible to individuals who currently do not have access to care. Passive monitoring could also be used as an adjunct to clinician-led treatment, thus increasing the quality of care and personalizing treatments.

Acknowledgments

This research was supported, in part, by Department of Homeland Security Award HSHQDC-16-3-00083, NSF CISE Expeditions Award CCF-1139158, Department of Energy Award SN10040 DE-SC0012463, Defense Advanced Research Projects Agency (DARPA) XData Award FA8750-12-2-0331, and gifts from Amazon Web Services, Google, IBM, SAP, The Thomas and Stacey Siebel Foundation, Apple Inc, Arimo, Blue Goji, Bosch, Cisco, Cray, Cloudera, Ericsson, Facebook, Fujitsu, HP, Huawei, Intel, Microsoft, Mitre, Pivotal, Samsung, Schlumberger, Splunk, State Farm, and VMware. Further support was provided by the Berkeley Institute of Data Science (BIDS). BR is generously funded by National Science Foundation award CCF-1359814, Office of Naval Research awards N00014-14-1-0024 and N00014-17-1-2191, the DARPA Fundamental Limits of Learning (Fun LoL) Program, a Sloan Research Fellowship, and a Google Faculty Award.

Conflicts of Interest

None declared.

References

1. Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998 Feb;66(1):7-18. [Medline: 9489259]
2. Kazdin AE. Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. Am Psychol 2008 Apr;63(3):146-159. [doi: 10.1037/0003-066X.63.3.146] [Medline: 18377105]
3. Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. Psychol Assess 1999;11(4):415-425. [doi: 10.1037/1040-3590.11.4.415]
4. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]

5. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
6. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
7. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
8. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Presented at: Proc of the ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka p. 1293-1304.
9. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
10. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. 2015 Presented at: International Conference on Affective Computing and Intelligent Interaction (ACII); 2015; Xi'an p. 222-228.
11. Gruenerbl A, Osmani V, Bahle G, Carrasco J, Oehler S, Mayora O, et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. 2014 Presented at: Proc of the 5th Augmented Human International Conference; 2014; Kobe.
12. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
13. Wang R, Aung MS, Abdullah S, Brian R, Campbell AT, Choudhury T, et al. Crosscheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. 2016 Presented at: Proc of the ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2016; Heidelberg p. 886-897.
14. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
15. LiKamWa R, Liu Y, Lane ND, Zhong L. Moodscope: building a mood sensor from smartphone usage patterns. 2013 Presented at: Proc of the 11th Annual International Conference on Mobile Systems, Applications, and Services; 2013; Taipei p. 389-402.
16. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]
17. Rabbi M, Ali S, Choudhury T, Berke E. Passive and in-situ assessment of mental and physical well-being using mobile sensors. Proc ACM Int Conf Ubiquitous Comput 2011;2011:385-394 [FREE Full text] [doi: 10.1145/2030112.2030164] [Medline: 25285324]
18. Healey J, Nachman L, Subramanian S, Shahabdeen J, Morris M. Out of the lab and into the fray: towards modeling emotion in everyday life. 2010 Presented at: Proc of the 8th International Conference on Pervasive Computing; 2010; Helsinki p. 156-173.
19. Byrne R, Eslambolchilar P, Crossan A. Health monitoring using gait phase effects. 2010 Presented at: Proc of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments; 2010; Samos.
20. Dimeo F, Bauer M, Varahram I, Proest G, Halter U. Benefits from aerobic exercise in patients with major depression: a pilot study. Br J Sports Med 2001 Apr;35(2):114-117 [FREE Full text] [Medline: 11273973]
21. Fox KR. The influence of physical activity on mental well-being. Public Health Nutr 1999 Sep;2(3A):411-418. [Medline: 10610081]
22. McKercher CM, Schmidt MD, Sanderson KA, Patton GC, Dwyer T, Venn AJ. Physical activity and depression in young adults. Am J Prev Med 2009 Feb;36(2):161-164. [doi: 10.1016/j.amepre.2008.09.036] [Medline: 19062235]
23. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry 2005 Mar;18(2):189-193. [Medline: 16639173]
24. Ströhle A. Physical activity, exercise, depression and anxiety disorders. J Neural Transm (Vienna) 2009 Jun;116(6):777-784. [doi: 10.1007/s00702-008-0092-x] [Medline: 18726137]
25. Pilcher JJ, Ott ES. The relationships between sleep and measures of health and well-being in college students: a repeated measures approach. Behav Med 1998;23(4):170-178. [doi: 10.1080/08964289809596373] [Medline: 9494694]
26. Pilcher JJ, Ginter DR, Sadowsky B. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. J Psychosom Res 1997 Jun;42(6):583-596. [Medline: 9226606]
27. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. 2013 Presented at: Proc of the 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); 2013; Cancun p. 145-152.

28. Incel OD, Kose M, Ersoy C. A review and taxonomy of activity recognition on mobile phones. Bionanoscience 2013;3(2):145-171. [doi: 10.1007/s12668-013-0088-3]
29. Lara O, Labrador M. A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tut 2012;15(3):1192-1209. [doi: 10.1109/SURV.2012.110112.00192]
30. Lu H, Yang J, Liu Z, Lane N, Choudhury T, Campbell A. The jigsaw continuous sensing engine for mobile phone applications. 2010 Presented at: Proc of the 8th ACM Conference on Embedded Networked Sensor Systems; 2010; Zurich p. 71-84.
31. Aharony N, Pan W, Ip C, Khayal I, Pentland A. Social fmri: investigating and shaping social mechanisms in the real world. Pervasive Mob Comput 2011;7(6):643-659. [doi: 10.1016/j.pmcj.2011.09.004]
32. Mizell D. Using gravity to estimate accelerometer orientation. 2003 Presented at: Proc of the 7th IEEE International Symposium on Wearable Computers (ISWC); 2003; White Plains p. 252.
33. Raudenbush SW, Bryk AS. Hierarchical linear models: Applications and data analysis methods. Thousand Oaks: Sage; 2002.
34. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 2013 Apr;68(3) [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
35. R Core Team. R-project. 2013. R: a language and environment for statistical computing URL: https://www.r-project.org/ [accessed 2017-09-07] [WebCite Cache ID 6tIUaTbQB]
36. Bates D, Maechler M, Bolker B, Walker S. cran.r-project. 2014. lme4: Linear mixed-effects models using eigen and s4 URL: https://cran.r-project.org/ [accessed 2017-09-07] [WebCite Cache ID 6tIVbxo1m]
37. Kuznetsova A, Brockho P, Christensen R. cran.r-project. 2014. lmerTest: Tests in Linear Mixed Effects Models URL: https://cran.r-project.org/ [accessed 2017-09-07] [WebCite Cache ID 6tIVkhOw4]
38. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20(3):273-297. [doi: 10.1023/A:1022627411411]
39. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:101093340]
40. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 2005;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
41. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830.
42. DeMasi O, Kording K, Recht B. Meaningless comparisons lead to false optimism in medical machine learning. CoRR 2017;abs/1707.06289 [FREE Full text]
43. Kempthorne O, Doerfler TE. The behaviour of some significance tests under experimental randomization. Biometrika 1969;56(2):231-248. [doi: 10.1093/biomet/56.2.231]
44. Smucker M, Allan J, Carterette B. A comparison of statistical significance tests for information retrieval evaluation. 2007 Presented at: Proc of the 16th ACM Conference on Information and Knowledge Management; 2007; Lisbon p. 623-632.
45. Aguilera A, Schueller SM, Leykin Y. Daily mood ratings via text message as a proxy for clinic based depression assessment. J Affect Disord 2015 Apr 1;175:471-474 [FREE Full text] [doi: 10.1016/j.jad.2015.01.033] [Medline: 25679202]
46. Teychenne M, Ball K, Salmon J. Physical activity and likelihood of depression in adults: a review. Prev Med 2008 May;46(5):397-411. [doi: 10.1016/j.ypmed.2008.01.009] [Medline: 18289655]

Abbreviations

BDI: Beck's Depression Inventory
BIDS: Berkeley Institute of Data Science
DARPA: Defense Advanced Research Projects Agency
GPS: global positioning system
mHealth: mobile health
RMSE: root-mean-square error
SMS: short message service
SVM: support vector machines

Edited by G Eysenbach; submitted 04.04.17; peer-reviewed by S Saeb, H Jin; comments to author 04.05.17; revised version received 29.06.17; accepted 28.07.17; published 05.10.17.

Please cite as:
DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B
Well-Being Tracking via Smartphone-Measured Activity and Sleep: Cohort Study
JMIR Mhealth Uhealth 2017;5(10):e137
URL: https://mhealth.jmir.org/2017/10/e137/
doi:10.2196/mhealth.7820
PMID:28982643

Original Paper

Determinants for Sustained Use of an Activity Tracker: Observational Study

Sander Hermsen[1], MSc; Jonas Moons[1], MSc; Peter Kerkhof[2], PhD; Carina Wiekens[3], PhD; Martijn De Groot[4], PhD

[1]Institute for Communication, Research Group Crossmedial Communication in the Public Domain, Utrecht University of Applied Sciences, Utrecht, Netherlands

[2]Department of Communication Science, Vrije Universiteit, Amsterdam, Netherlands

[3]Centre of Expertise Energy, Hanze University of Applied Sciences, Groningen, Netherlands

[4]Quantified Self Institute, Hanze University of Applied Sciences, Groningen, Netherlands

Corresponding Author:
Sander Hermsen, MSc
Institute for Communication
Research Group Crossmedial Communication in the Public Domain
Utrecht University of Applied Sciences
Bolognalaan 101
Utrecht, 3584 CJ
Netherlands
Phone: 31 884813953
Email: sander.hermsen@hu.nl

Abstract

Background: A lack of physical activity is considered to cause 6% of deaths globally. Feedback from wearables such as activity trackers has the potential to encourage daily physical activity. To date, little research is available on the natural development of adherence to activity trackers or on potential factors that predict which users manage to keep using their activity tracker during the first year (and thereby increasing the chance of healthy behavior change) and which users discontinue using their trackers after a short time.

Objective: The aim of this study was to identify the determinants for sustained use in the first year after purchase. Specifically, we look at the relative importance of demographic and socioeconomic, psychological, health-related, goal-related, technological, user experience–related, and social predictors of feedback device use. Furthermore, this study tests the effect of these predictors on physical activity.

Methods: A total of 711 participants from four urban areas in France received an activity tracker (Fitbit Zip) and gave permission to use their logged data. Participants filled out three Web-based questionnaires: at start, after 98 days, and after 232 days to measure the aforementioned determinants. Furthermore, for each participant, we collected activity data tracked by their Fitbit tracker for 320 days. We determined the relative importance of all included predictors by using Random Forest, a machine learning analysis technique.

Results: The data showed a slow exponential decay in Fitbit use, with 73.9% (526/711) of participants still tracking after 100 days and 16.0% (114/711) of participants tracking after 320 days. On average, participants used the tracker for 129 days. Most important reasons to quit tracking were technical issues such as empty batteries and broken trackers or lost trackers (21.5% of all Q3 respondents, 130/601). Random Forest analysis of predictors revealed that the most influential determinants were age, user experience–related factors, mobile phone type, household type, perceived effect of the Fitbit tracker, and goal-related factors. We explore the role of those predictors that show meaningful differences in the number of days the tracker was worn.

Conclusions: This study offers an overview of the natural development of the use of an activity tracker, as well as the relative importance of a range of determinants from literature. Decay is exponential but slower than may be expected from existing literature. Many factors have a small contribution to sustained use. The most important determinants are technical condition, age, user experience, and goal-related factors. This finding suggests that activity tracking is potentially beneficial for a broad range of target groups, but more attention should be paid to technical and user experience–related aspects of activity trackers.

(JMIR Mhealth Uhealth 2017;5(10):e164) doi:10.2196/mhealth.7311

KEYWORDS

mobile health; mHealth; physical activity; machine learning; habits

Introduction

The Effect of Activity Tracker Usage on Physical Activity

One of the biggest threats to our health is physical inactivity, which is considered to cause 6% of deaths globally [1]. Too little physical activity plays a role in a range of debilitating conditions such as cardiovascular diseases, diabetes mellitus type II, chronic obstructive pulmonary disease, and some forms of cancer [2,3]. The American Heart Association endorses 10,000 steps a day or 30 min of moderate-intensity physical activity (eg, brisk walking) for at least 5 days a week as guidelines to improve health and reduce health risk [3,4]. Unfortunately, many people fail to meet these criteria [5].

Behavior change toward more physical activity might greatly benefit our health. Unfortunately, for many people, their physical activity is a deeply engrained habit [6,7]. Choosing physical activity over inactivity tends to occur outside awareness [7]. This lack of conscious scrutiny is one of the main reasons sedentary habits are difficult to change; we are not always adept in monitoring our own behavior, especially not when this behavior is executed unintentionally [8]. It is not surprising, therefore, that people tend to overestimate their physical activity [9,10]. Supporting our self-monitoring abilities by providing us with timely and relevant feedback on our behavior has proven a successful strategy to disrupt the automaticity of deeply engrained habitual behaviors such as inactivity and make them available for conscious scrutiny [11-13].

In recent years, numerous interactive and mobile technology solutions to encourage physical activity have arrived in the form of devices that are able to directly monitor our physical activity through a range of sensors. The information thus gathered can be applied by automatically providing the user of the device with behavior change techniques (BCTs) from the monitoring cluster [14]: timely feedback on their own behavior and the possibility to self-monitor behavior and its outcomes. Furthermore, dashboard applications often encourage (but hardly ever enforce) a range of secondary BCTs: goal setting, the review of behavioral goals and their outcomes, and social comparison and support.

Such activity trackers are an increasingly popular way to promote physical activity. In 2012, a survey showed that 69% of adults in the United States tracked at least one health behavior using some sort of tracking device, and 14% of US citizens owned a specialized activity tracker of some sort [15]. Of those who did track a health behavior, roughly half indicated that tracking changed their overall approach to maintaining their health (ibidem).

The effect of using activity tracker technology on physical activity is well established for a range of populations (eg, [16-19]); however, a crucial ingredient for lasting effects of behavior change interventions in general is the sustained use of the intervention [12]. Unfortunately, even though there is a growing body of research utilizing activity trackers, there is as yet little research available on sustained use of such devices. Anecdotal evidence, as well as what little evidence that is available [19], suggests activity trackers may have a poor record when it comes to sustained use, as they are easy to switch off, ignore, lose, or neglect. Furthermore, there is to date no research available that sheds light on which users manage to stick to using their activity tracker during the first year (and thereby increasing the chance of healthy behavior change) and which users stop using their trackers after a relatively short time. This paper attempts to add to our knowledge of the sustained use of activity trackers and factors that predict this sustained use.

Potential Determinants of Tracker Use

On the basis of evidence from prior research on the effect of feedback interventions on habitual behaviors (eg, [12,20]), there is a broad range of factors that might influence sustained use and efficacy of activity trackers.

First, tracker *technology* may play a crucial role in sustained use. Trackers may be abandoned because of empty batteries, with the perceived cost of replacement too high or too cumbersome [21]. Apart from technical failures, actual or perceived characteristics of the tracker may fit user expectations. The user experience and ease of use [22,23], functionality or lack thereof [22], the possibility to upgrade toward a newer device (ibidem), aesthetics and form [24], perceived accuracy (ibidem), and perceived fit between device and self-image [23] are all reasons to either abandon the tracker or to keep using it. Furthermore, the data delivered by the tracker must fit participants' needs (ibidem). Finally, computer literacy, or the perceived self-efficacy in using digital devices, is known to affect sustained use (eg, [25,26], higher more than lower).

Socioeconomic status markers such as *education* (eg, [25,27], higher more than lower) and *employment* (eg, [28,29], higher more than lower), *age* [25,30,31], older more than younger) and *gender* (eg, [25,27,32], women more than men) are known to influence sustained use, as are *psychological traits* such as inhibitory strength and the capacity for self-regulation [33-35].

Personal *health-related factors* may very well influence the sustained use of the activity tracker; poor health decreases perceived self-efficacy [36,37], which is known to influence sustained use [38]. Low mood, stress, sleep disturbances, and other markers of mental health, are also known to decrease sustained use [30,39].

Goal-setting is generally seen as a promising strategy to increase the use of physical activity interventions [40-42]. Strong, clear goals and *motivation* to fulfil these goals [25,30,31] increase the chance of sustained tracker use. *Achieving* these goals, or at least displaying a performance level that could lead to achieving previously-set goals, can provide a further boost to initial motivation and perceived self-efficacy, increasing the chances of sustained tracker use. However, the fulfilment of a set goal may also lead to device abandonment, because users feel they no longer need the tracker [43].

Furthermore, behavior change theories (eg, social cognitive theory [44] and control theory [35]) suggest that behavior change is most likely if feedback is not delivered on its own but *embedded* in larger interventions with clear target behaviors and action plans. Combined use of the activity tracker with other health apps, participation in a therapeutic regime, and use of the app and Web-based platform that accompany the activity tracker may be seen as an operationalization of this concept of integration. Overall, we expect users with strong goals and high integration of their tracking behavior in other health-related practices to have a higher chance of sustained tracker use, especially when these users manage to achieve their performance goals.

Feedback properties such as timing, duration, frequency and sensory modality (cf [20]), and *user experience* (eg, [45]) are known to influence the efficacy of the feedback intervention, both directly and through perceived usability and agreeableness. Similarly, feedback properties [46] and user experience-related factors are known to affect the uptake and sustained use of physical activity trackers [47]. We expect users with greater liking of the tracker and its accompanying online tools to have a higher chance of sustained tracker use.

Activity tracking is often social and collaborative instead of individual and personal [24,48,49]. *Social interaction* is known to improve adherence to physical activity interventions in general [50]. We therefore expect users that share their tracking data with peers or relatives to have a higher chance of sustained tracker use.

Sample Size and Duration in Previous Research on Activity Trackers

Current research into determinants of activity tracker use typically makes use of small test populations, ranging from 7 to 31 participants (eg, [23,24,48,49,51-53]), which limits the possibilities to reliably investigate quantitative measures of determinants of device use. When larger samples have been tested (eg, [22], n=1561 and [39], n=256), only a small number of determinants were included. Furthermore, adherence studies generally covered only a very short period, that is, 2 months or less (eg, [23,51-53]). Only one study ([54]) tested sustained use over a period of up to 10 months. However, this study did not evaluate potential determinants for adherence.

This study attempts to contribute to bridging this knowledge gap by looking into factors predicting sustained use in the first year after purchase. Specifically, we look at demographic and socioeconomic, psychological, health-related, goal-related, technological, user experience–related, and social predictors of feedback device use and their predictive power in determining which participant is most likely to continue using the device.

Methods

Study Design

This study was initiated by IDS Santé Inc (Paris, France), a full-service communication agency aimed at the health sector and specializing in prevention and health education and executed from June 2013 until winter 2014 as a project called "MySantéMobile." A total number of 1000 participants were recruited in France via a (free) newspaper from four French cities (Bordeaux, Lille, Montpellier, and Lyon). Each participant received an activity tracker and was requested by email to fill in three Web-based questionnaires (June 2013, August 2013, and January 2014). After completion of the study, the full raw dataset was transferred for independent and retrospective analysis to the authors of this paper.

To establish which set of the included predictors best explains the use and nonuse of this activity tracker in the dataset, we adopted the Random Forest method, a machine learning approach [55]. This approach enables identification of predictors that explain large portions of variance while minimizing the risk of overfitting, which is likely to occur when performing a regression analysis with a large set of predictors [56]. Furthermore, this approach is also capable of detecting nonlinear relationships and higher-order interactions between predictors.

Activity Tracker

The activity monitor used in this study, the Fitbit Zip, is a small (2.9 cm x 3.6 cm x 1 cm) consumer device that tracks activity through counting steps. The Zip is worn as a clip-on device on the waist or elsewhere where it can be easily clipped onto clothing. On the device screen, the Zip displays the number of steps taken on the current day, and, after pressing a button on the device, displays the distance covered on the current day, active minutes, the time, an approximation of calorie expenditure, and feedback in the form of a happy, neutral, or unhappy *smiley*. Research [57,58] shows that the reliability and validity of the Fitbit Zip activity monitor is high, with little error in the number of registered steps, both in laboratory conditions and in daily life.

Participants

Recruitment

Participants were recruited through a newspaper article, published on the 14th of May 2013, in free newspapers in France. 1000 participants were selected using the following inclusion criteria: living in one of the four eligible cities (Montpellier, Lyon, Lille, and Bordeaux); at least 18 years of age; and owning a smartphone or computer compatible with Fitbit. Of those 1000, 929 received a Fitbit Zip activity tracker and took part in the study.

Data Acquisition

In the first week of June (2013), all eligible participants were invited to fill out a Web-based questionnaire by email. This questionnaire was presented through the LimeSurvey platform and covered sociodemographics, device usage, tablet/phone brand, self-reported tracker use, use of other health apps and devices, health, exercise, and diet. All questionnaires used in this study are available in Multimedia Appendix 1. Approximately two weeks after filling in this questionnaire, the participant received their Fitbit Zip tracker by mail. Participants received their Fitbit Zip tracker free of charge.

Upon dispatch of the Fitbit trackers, participants received an email giving them instructions on how to install and use the Fitbit, how to synchronize data and how to authorize

MySantéMobile in acquiring their data through the Fitbit API. Instructions were also provided on the MySantéMobile website. Participants then had to give permission to MySantéMobile to read their activity data through the Fitbit API. Participants who did not give permission received reminder phone calls and emails.

A second questionnaire was sent out by email on 23 August 2013 (after 98 days). The third questionnaire was also sent out by email, on 7 January 2014 (232 days). Participants who did not fill out the questionnaire received a reminder email after two weeks. Participants received no incentive other than a free activity tracker. At the end of the data acquisition period, all participants received an overview of the study results.

Participant Selection

Since the selected analysis method does not allow missing values, only data from those participants who completed their questionnaires could be used. Of the 929 participants originally approached to take part in the study, 711 participants (76.5%) completed the first questionnaire and gave permission to MySantéMobile to read their activity data through the Fitbit API (Textbox 1). Data collection using the Fitbit took place from 20 June 2013 to 13 May 2014 (327 days). Of this group of 711 participants, a total number of 575 participants (80.8%) completed the second questionnaire (August 2013) and 542 participants (76.2%) completed both the second and the final questionnaire (January 2014).

Textbox 1. Participant characteristics at Q1.

- Gender
  - 330 female, 381 male
- Age
  - < 25: 133, 26-35: 444, 36-45: 182, 46-55: 124, 56-65: 49, >65: 4
- Marital status
  - Single: 240, Couple: 332, Single parent: 38, Family: 272, Other: 54
- Profession
  - Cadre (management): 456, Intermédiaire (middle management): 91, Employé (employee): 251, Artisan (craftsperson): 51, Ouvrier (worker): 12, Retraité (retired): 19, Sans (without): 56
- Education
  - Bac: 106, Bac+2: 371, Bac+5: 406, CAP/BEP: 38, Brevet des Colleges: 11, None: 2

Measures

The total number of days on which the device was worn was used as the primary outcome measure (adherence to using the wearable for self-tracking). We only had access to data that were synchronized with a personal computer or mobile app. However, the Fitbit Zip stores steps data for 30 days, therefore, we assume most active users will synchronize their data within this time window. For ease of interpretation, we will speak of "using" or "wearing" the Fitbit. However, note that our measure may somewhat underestimate the number of days the Fitbit was worn.

Furthermore, we calculated the average amount of steps taken by each participant on those days the tracker was used.

Questionnaires and Item Selection

Three questionnaires (Q1, June 2013; Q2, August 2013; Q3, January 2014) were sent out to the participants. A complete overview of all three questionnaires, with the exact questions (translated into English), and the response scales used for each question, is available as Multimedia Appendix 1.

From these questionnaires, we selected for our analysis those items that (1) matched the potential determinants for sustained use of the tracker outlined in the introduction of this paper, and (2) met with our requirements for item validity.

On the basis of our analysis of potential determinants for sustained use, we included the following items from the questionnaires in our analysis:

1. Demographical and socioeconomic factors: age, gender, place of residence, household size and household composition, profession, and education (all in questionnaire 1 (Q1).
2. Psychological factors: general mood (all questionnaires); specific scores on affective situation (sadness, gaiety), stress (calmness, stressfulness), energy (energy level, tiredness), and sleep quality (all in all questionnaires); big five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism; plus, rebelliousness, health-mindedness, and independence (all in Q3).
3. Technological factors: synching platform type (smartphone, tablet, computer), operating system—iOS or Android (all in Q1), use of other health applications (Q1), experience with technology (Q3).
4. User experience: perceived utility, enjoyableness, intrusiveness, modernity, fun, reliability, simplicity, inconvenience, correspondence to needs, beauty, robustness, and cumbersomeness of the activity tracker (all in Q2); exactness, detail, clarity, credibility, confidence, insight, perceived efficacy (all in Q3).
5. Health-related factors: body mass index (all questionnaires), smoking (Q1), pregnancy (Q1), diet (Q1), medical treatment

status (Q1), activity in sports (Q1), and sports together with others (Q1).

6. Predefined participant goals and perceived goal achievement: increasing activity, improving sleep, quitting smoking, diagnosing or improving diet, diagnosing behaviors, losing weight, and improving stamina (all in Q1); for each goal, the perceived achievement of the goal was measured (Q2 and Q3).
7. Social factors: whether participants talked about the tracker sharing use with family, friends, colleagues, teams and clubs; sharing data on the Internet through social media, blogs, Twitter, websites, forums, and mailing lists (Q2 and Q3).

Questionnaire Validity

Because of the history of this study, which started as groundwork for a publicity campaign for a communications agency, the questionnaires used in this study have not been constructed in such a way that meets the current standards for validity. To evaluate the validity of the three questionnaires used in this study and to determine which items were of high enough standard to include in our analysis, we compared each question with current, well-validated standard approaches in scientific literature. The complete result of this analysis is included in Multimedia Appendix 1. For each item, under "remarks," the validity evaluation is listed. Generally, our evaluation showed that the greater part of the questionnaire items survives rigid scrutiny and satisfies scientific criteria. However, the validity of four items, one item on digital proficiency and three items on psychological traits (rebelliousness, independence, and health-mindedness) could not be satisfactorily assessed. Results for these items should be used with caution.

The greater part of the questionnaire consisted of single-item measures. Single-item measures can be eminently usable (sometimes even more so than multiple item measures) when the attribute (eg, attitude, frequency) is concrete and singular (ie, not consist of multiple facets) and when the object of the item (eg, brand, product) is concrete [59]. For most of the items, this is the case; see Multimedia Appendix 1 for an overview. Three exceptions occurred, which are as follows: items regarding emotional well-being, items regarding psychological traits, and items regarding user experience.

The first group of items that do not have a concrete object are about emotional well-being. However, single-item assessments of emotional well-being are often used in large-scale surveys (see [60] for an overview) and have been shown to perform quite well compared with multiple-item scales (eg, [61]). We can therefore probably conclude that these single-item self-report measures are a sufficiently valid measure for the purpose of this paper.

The second group of items that do not have a concrete object concern psychological traits. In Questionnaire 3, the big five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) are measured using the French translation of the TIPI questionnaire [62]. This questionnaire has been well validated.

A third group of items address the user experience of the Fitbit. User experience can be defined as *a person's perceptions and responses that result from the use and/or anticipated use of a product, system, or service* [63]. In recent years, approximately a hundred different measures have been developed [64]. User experience evaluations generally consist of questions addressing some (but never all) of the following concepts: timeliness, adaptability, comfort, opacity, efficiency, immersion, intuitiveness, ease of use, usefulness, interaction, controllability, clearness, completeness, identity, novelty, originality, fun, stimulation, valence, connectedness, attractiveness, beauty, and trust [65]. Currently, no questionnaires or other measures exist that address all of the aforementioned concepts. In this study, comfort, ease of use, novelty, fun, valence, attractiveness, trust, and invasiveness have been measured.

Concepts regarding user experience can be subdivided in three categories [66]: pragmatic qualities (usability-oriented; first 13 concepts [timeliness-completeness]), hedonic qualities (identity-connectedness), and general concepts, mainly focused on attraction. Principal component analysis showed that the response patterns for the 22 user experiences-related items asked in questionnaires 2 and 3 justified the construction of three conceptual factors. The first factor was the valence of the activity tracker, which was formed by the following 12 items: usefulness (practicality), niceness, modernity, amusingness, credibility, ease of use, level of answering to needs, beauty, robustness, intrusiveness, embarrassment, and nuisance. This factor corresponds with the hedonic quality in [66]. The second factor was the preciseness of the activity tracker, which was formed by the following 4 items: exactness, level of detail, clarity, and credibility. This factor corresponds with the pragmatic qualities in [66]. The third and final factor was perceived efficacy of the activity tracker, which was constructed by averaging 3 items on perceived efficacy of the tracker, namely activity increase, health changes, and well-being. This third factor does not correspond with hedonic nor pragmatic qualities but has to do with the perceived effectiveness of the activity tracker. The results of the principal component analysis are reported in Multimedia Appendix 2.

On the basis of our analysis, some items were left out of our analysis; see Multimedia Appendix 1 under "left out" for each questionnaire. Four items, of which the validity could not be determined satisfactorily, were nevertheless included in the analysis. One item concerned technological aptitude, and three further items concerned the psychological traits: rebellion, independence, and health-mindedness. We advise to treat the results of these items with caution.

Statistical Analysis

We determined the relative importance of all included predictors by using Random Forest, an analysis technique based on recursive partitioning [55,56]. Random Forest is an ensemble method that makes use of a large number of decision trees, strengthened by "bootstrap aggregating": drawing random samples from the original dataset with replacement. For each of the bootstrap samples that are drawn, a decision tree is constructed. At each branch of the tree, a random selection of the predictor variables is considered. The variable that produces

the best split (ie, most informative and offering the largest contrast) is used to divide the cases over two daughter nodes.

To predict the outcome variable for a specific case, Random Forest uses the predictions of all trees to arrive at an "ensemble" prediction (in the case of regression, it averages the prediction of all trees). To evaluate the performance of the Random Forest model, we can test each tree on those cases that fell outside its bootstrapped sample and thus, were not used to grow the tree. This produces an "out-of-bag" error rate, which is a good approximation of the test error.

Random Forest analysis can produce a list of predictors, sorted by relative importance. This is done by calculating the mean squared error (MSE) and looking at the relative increase of the MSE when the values of a predictor are permuted across cases. Permuting the predictors retains frequency information but destroys the association between the predictor and the outcome variable. If the variable is important for the Random Forest model, we would expect its predictions to deteriorate and the MSE to go up. Thus, the relative increase in MSE is used to determine an importance ranking of predictors, sorted from greatest to least increase in MSE.

Random Forest modeling has the benefit of being able to deal with large numbers of predictor variables with complex interactions, especially in situations with relatively few cases relative to the number of predictors. Furthermore, Random Forest is capable of detecting nonlinear relations between independent and dependent variables. Random Forest analysis methods have recently been applied successfully in genetics, clinical medicine, bioinformatics, and the social sciences (see [56,67] for examples).

The predictor variables for the Random Forest analysis were taken from the responses to the questionnaires. All parameter settings, source code, and data files for the Random Forest analysis are available through the Open Science Foundation.

We used R 3.3 for analysis [68] and the R package "randomForest" [69] for Random Forest modeling.

Results

Fitbit Use

The mean number of days that participants used their Fitbits was 129.3 (nonconsecutive) days (standard deviation [SD]=88.5; median=122). Figure 1 shows the distribution of total days of use.

A graphical overview of usage over time is shown in Figure 2. As some users only started to wear the Fitbit after weeks or even months, the number on the x-axis refers to the number of days since the first day the device was used, rather than from the start of the study. The figure indicates both the percentage of participants who used the device for any length of time after the indicated day and habitual use (defined as 3, 5, and 7 days worn out of the last 7 days). The decline during the first 50 days coincides for most users with the French holiday season (July-August). The peak shortly after 100 days coincides with Q2 being sent out (again, for most users).

The pattern of (nonhabitual) usage decline is roughly linear. A linear regression with time as the independent variable shows a decline of 2.0 percentage points per week from day 1 to day 300. In other words, every week, 2% (14) of the participants at start stopped using the tracker entirely. As the base of users is shrinking, this means that the proportion of participants who stopped using the tracker increases over time. After 175 days (5.7 months), 50% of users have stopped wearing their tracker.

Habitual use seems to follow a pattern of slow exponential decay. An exponential model shows that the proportion of users wearing the activity tracker 5 or more days per week declines 5.7% per week, as calculated from the peak after the summer holiday dip (from day 102 to day 300).

On average, participants took 7492 steps (SD 3012) per synced day (ie, day on which they wore their tracker). The median number of steps was 7107, with an interquartile range of 3462. A plot of the distribution of the number of steps taken is provided in Figure 3.

An overview of the correlations between the number of days on which the tracker was used, mean number of steps, and a range of self-report measures on personal health are displayed in Table 1. The number of days the activity tracker was used significantly predicted mean steps per day: b=9.43, t_{709}=32.60, and P<.001. A significant proportion of the variance was explained: R^2=.08, $F_{1,709}$=8.92, and P<.001. An exploratory analysis of correlations with measures on personal health showed generally weak to negligible associations for both days used and mean steps per day. Strongest associations were between the number of days used and self-reported general health (r=.16) and between days used and physical shape (r=.12).

Table 1. Correlations (Spearman *r*, unless marked with b: Pearson *r*) between steps taken, days worn, and health measures. All measures are from Q3 unless otherwise noted.

Health measure	Mean steps	Days used
Mean steps		.28[a,b]
Days used	.28[a,b]	
Self-reported weight change (kg, Q3-Q1)	−.05[a]	−.04[a]
Self-reported effect tracker on weight	.04	.10[d]
Self-reported effect tracker on activity	.02	.04
Self-reported effect tracker on sleep quality	.05	.08
Self-reported effect tracker on smoking	−.03	−.07
Self-reported effect tracker on healthy eating	.04	.07
Self-reported effect tracker on general health	.04	.16[b]
Self-reported effect tracker on physical shape	.07	.12[c]
Self-reported tiredness	−.04	−.06
Self-reported happiness	.04	.07
Self-reported stress	−.04	.01
Trend in self-reported tiredness (Q3-Q1)	.07	.03
Trend in self-reported happiness (Q3-Q1)	.04	−.01
Trend in self-reported stress (Q3-Q1)	.03	−.01

[a]Pearson *r*.

[b]Significance at *P*<.001.

[c]Significance at *P*<.01.

[d]Significance at *P*<.05.

Figure 1. Distribution of participants' total number of days of activity tracker use.

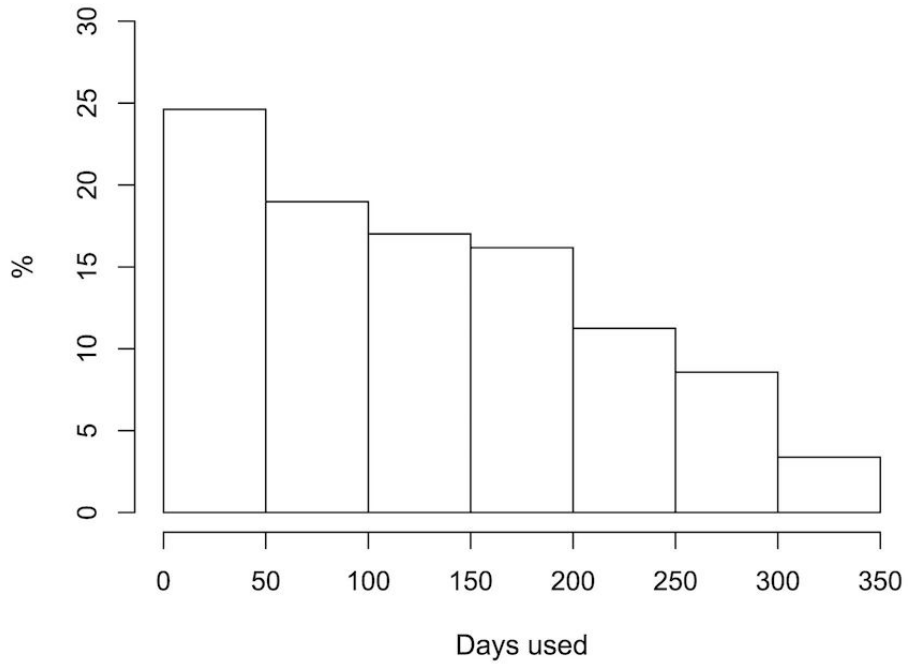


Figure 2. Usage decline over time. The horizontal axis shows the number of days since the first day of use. The percentage of participants who used the activity tracker for any number of days after a particular day is indicated with a solid line. The other lines indicate habitual use: the percentage of participants who used the tracker for at least 3, 5, and 7 days in the preceding 7 days. Note that this includes participants who stop using the tracker and later start using it again. The early dip in use is due to the summer holiday.

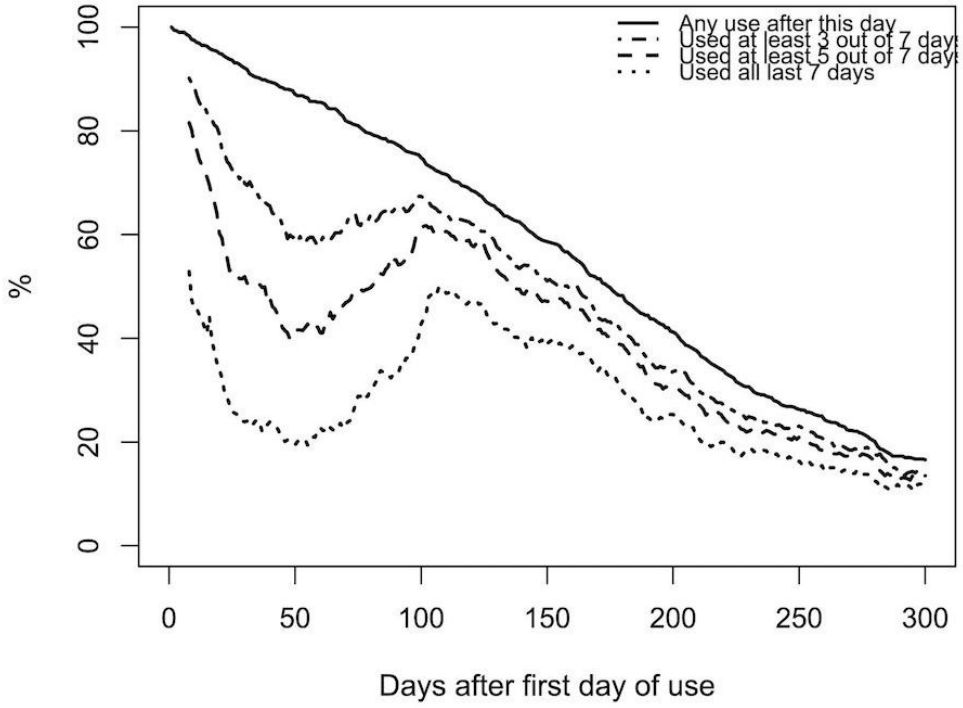


Figure 3. Distribution of participants' mean number of daily steps.

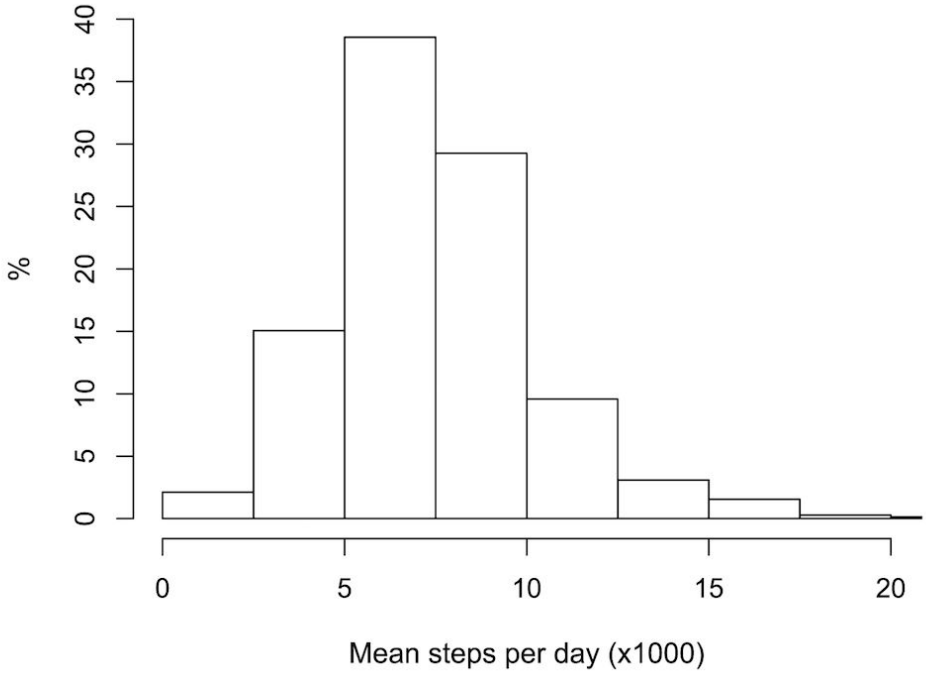


Table 2. Reasons for not wearing the Fitbit.

Reason to not wear	Q2 (98 days)			Q3 (232 days)		
	Count	% of total reasons	% of total respondents Q2 (639)	Count	% of total reasons	% of total respondents Q3 (601)
Technical failure or difficulty	20	50.0	3.1	106	56.7	17.5
Lost the device	9	22.5	1.4	24	12.8	4.0
Forgot to wear	2	5.0	0.3	24	12.8	4.0
Had no use for device or no motivation	0	0.0	0.0	16	8.6	2.6
Health issues	1	2.5	0.2	7	3.7	1.2
Used other device	0	0.0	0.0	2	1.1	0.3
Was on holiday	7	17.5	1.1	2	1.1	0.3
Did not start yet	1	2.5	0.2	0	0.0	0.0
Other	0	0.0	0.0	6	3.2	1.0
Total	40	100	6.3	187	100	30.8

Reasons for No Longer Using the Tracker

In both Q2 and Q3, participants were asked how many of the last 30 days they wore their Fitbit. If the answer was “fewer than 5,” they were asked additionally why they did not wear their tracker (more often) in an open-ended question. The responses were categorized and can be found in Table 2.

The results from Q2 (sent after 98 days) indicated that 40 participants (6.3% of all respondents) used the Fitbit fewer than 5 days in the last month. The primary reason for not using the device, given by half of those indicating low or nonuse, was technical failure or other technical problems, including empty batteries. Other reasons included losing the device or being on a holiday. Technical problems were also the main reason given in Q3 (sent after 232 days), with 17.5% of all respondents reporting this issue.

Factors Associated With Usage

We used the Random Forest method to investigate which predictors are associated with continued use of the activity tracker. The total number of days on which the device was worn was used as the outcome variable. As many participants did not respond to all three questionnaires, with those who stopped using their Fitbit less likely to fill in questionnaires 2 and 3, we decided to construct two different models, corresponding to two different groups of participants: (1) participants completing Q1 and (2) participants completing all three questionnaires (Q1 to Q3). For the latter model, we analyzed only those participants who did not state technical malfunction of any kind as a reason to quit.

In the first model, the data from those 586 participants who completed Q1, gave permission to use their tracker data, and stated neither technical issues nor lost trackers as a reason to no longer track, were entered. Some predictors were adjusted or recalculated. A complete overview of all questionnaires, the exact questions, the response scales, and any recalculations or adjustments is available in Multimedia Appendix 1.

Figure 4 shows the relative impact of each predictor variable in Q1 on the amount of variance explained, expressed as the relative increase in MSE when the predictor is randomly permuted across participants. The Random Forest model with all Q1 predictors included explains 8.29% of variance. Only those predictors whose increase in MSE is above zero are displayed because decreases can be safely attributed to noise.

We explored the effect of the different predictors on the number of days the participants wore their tracker. Figures 5-9 provide boxplot representations of the distributions for the marginal means (with all other factors kept constant) of the different levels of each predictor. Only those predictors of which the differences in marginal means implies a meaningful difference in real life (>1 day) are included: Age, goal to quit smoking, iPhone type, sports activities in the company of others, household type, household size, having a smartphone, having an iOS-based-smartphone, profession, and smoking.

In the second model, data from 397 participants who completed all three questionnaires (Q1 to Q3) and who did not state technological malfunction as a reason to stop tracking, were entered. Once again, some predictors were adjusted or recalculated. A complete overview of all questionnaires, the exact questions, the response scales, and any recalculations or adjustments is available in Multimedia Appendix 1.

Figure 6 shows the relative impact of each predictor variable in the questionnaires on the amount of variance explained, expressed as the relative increase in MSE when the predictor is randomly permuted across participants. The total percentage of variance explained by the Random Forest model with all Q1 to Q3 predictors is 10.91%.

Once again, we explored the effect of the different predictors on the number of days participants wore their tracker. Figures 11-15 show the marginal means (with all other factors kept constant) of the different levels of each predictor. Again, only those predictors of which the differences in marginal means implies a meaningful difference in real life (>1 day) are included: age, perceived effect on goals, user experience

(valence), user experience (effect), perceived effect on goals, iPhone type, and the goal to change eating habits. User experience (valence and effect) and perceived effect on goals are continuous variables, so no levels of marginal means could be shown. Instead, we show a partial dependence plot. The user experience variables are shown normalized with a mean 0 and a SD of 1.

Figure 4. Plot of relative importance of predictors of sustained use in questionnaire 1 (Q1); BMI: body mass index, MSE: mean squared error.

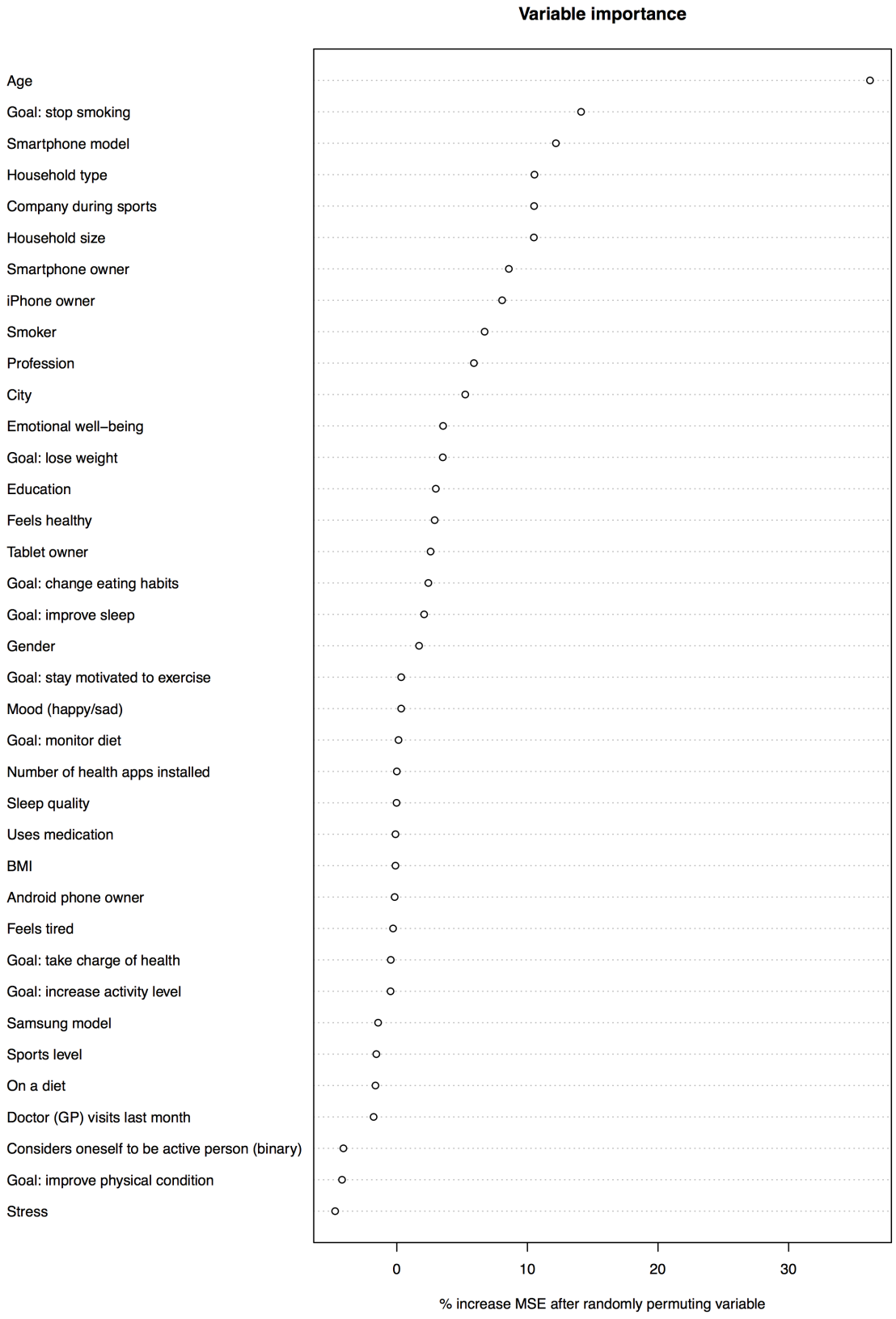


Figure 5. Boxplots of the distributions of Age levels. Older participants have longer sustained use.

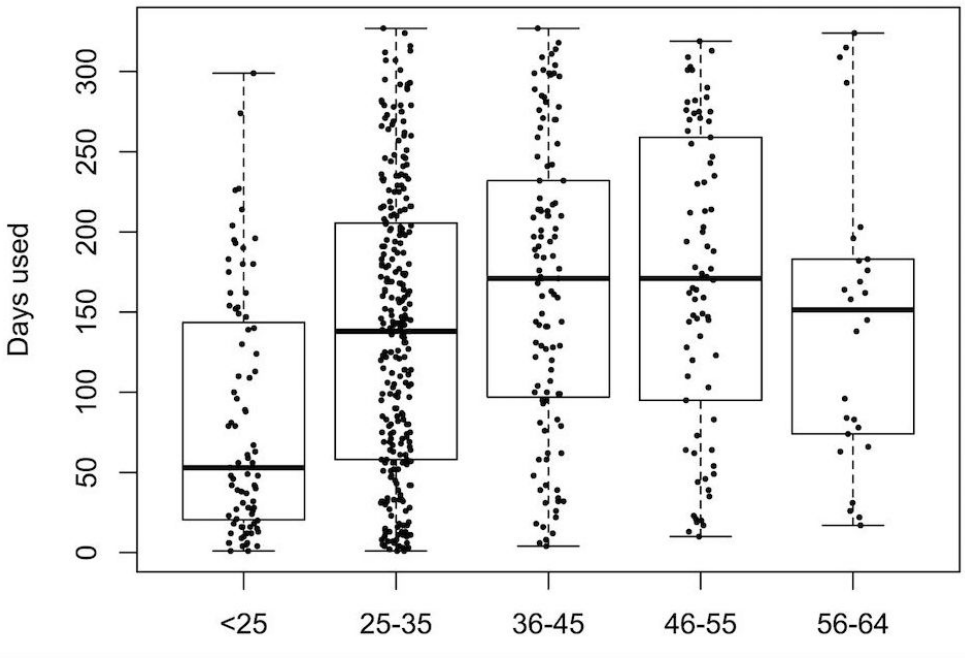


Figure 6. Boxplots of the distributions of iPhone type levels. Holders of iPhones show less sustained use than those of other smartphones.

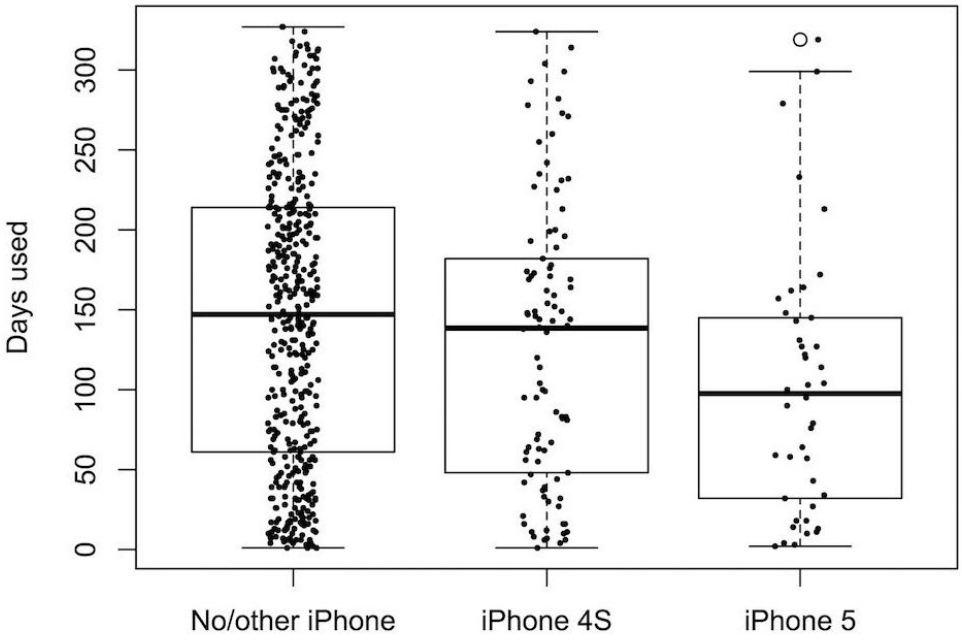


Figure 7. Boxplots of the distributions of Having the goal to quit smoking. Those not wanting to quit smoking (including non-smokers) have longer sustained use than those who do.

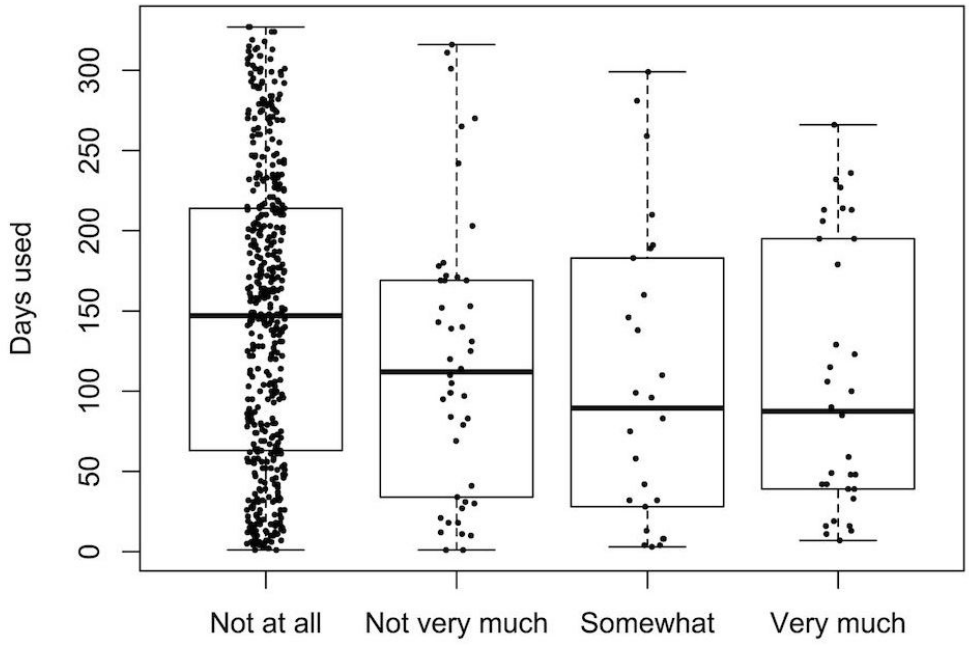


Figure 8. Boxplots of the distributions of Household type. Single parents show shorter sustained use than other household types.

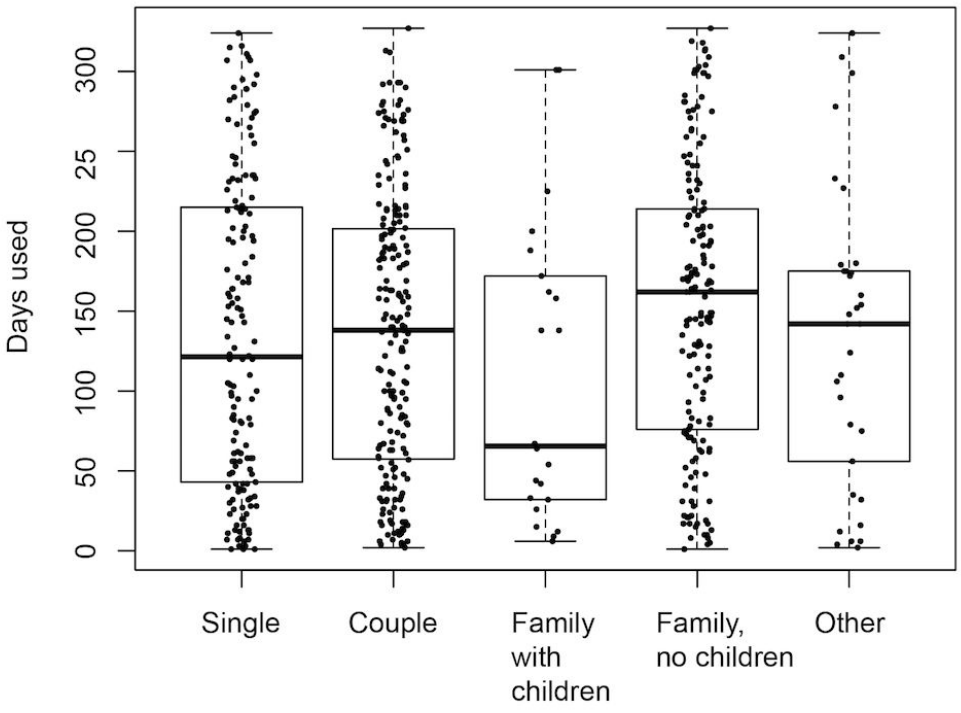


Figure 9. Boxplots of the distributions of Sports in company of others. Those who practice individual sports or with relatives, have longer sustained use of the tracker than those who participate in sports with friends or acquaintances.

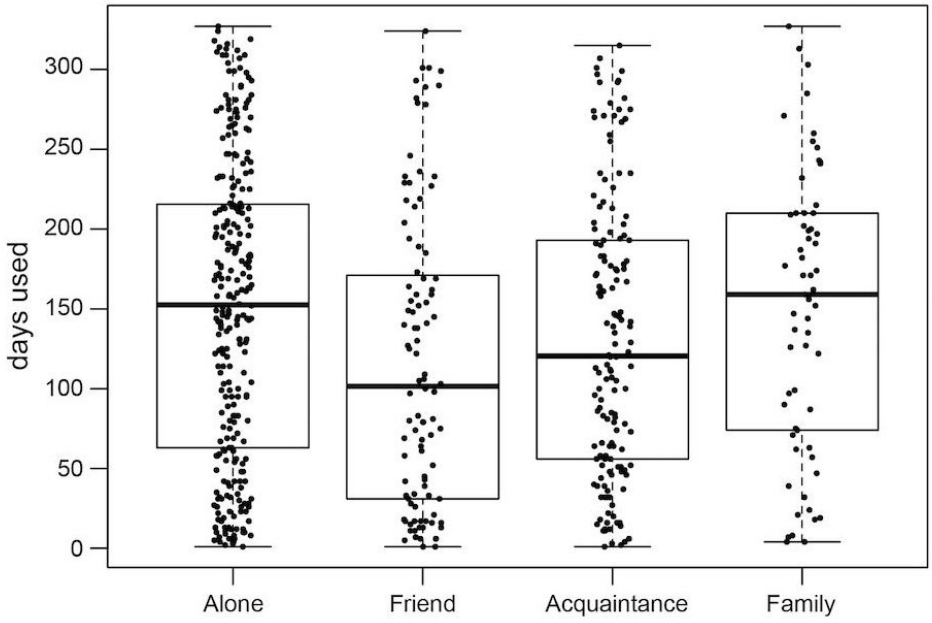


Figure 10. Plot of the relative importance of predictors in all questionnaires (Q1 + Q2 + Q3).

Figure 11. Boxplots of the distributions of the marginal means for participant age. The under-25 use the Fitbit less long than the other groups. There are no participants older than 65 in this sample.

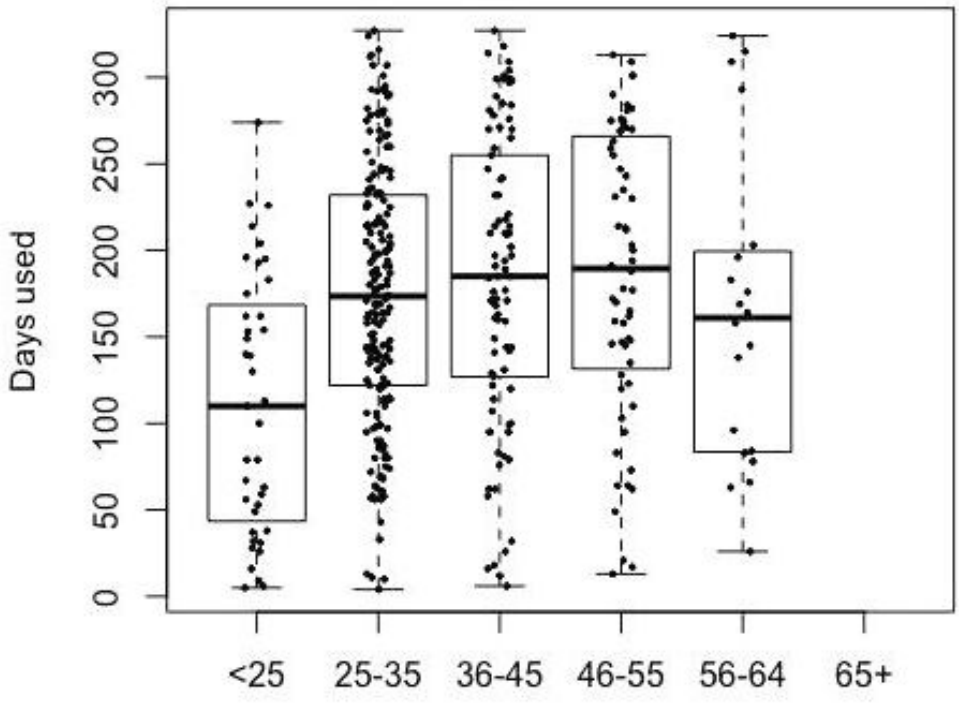


Figure 12. Boxplots of the distributions of having the goal to change eating habits. The stronger the goal, the less sustained use of the tracker.

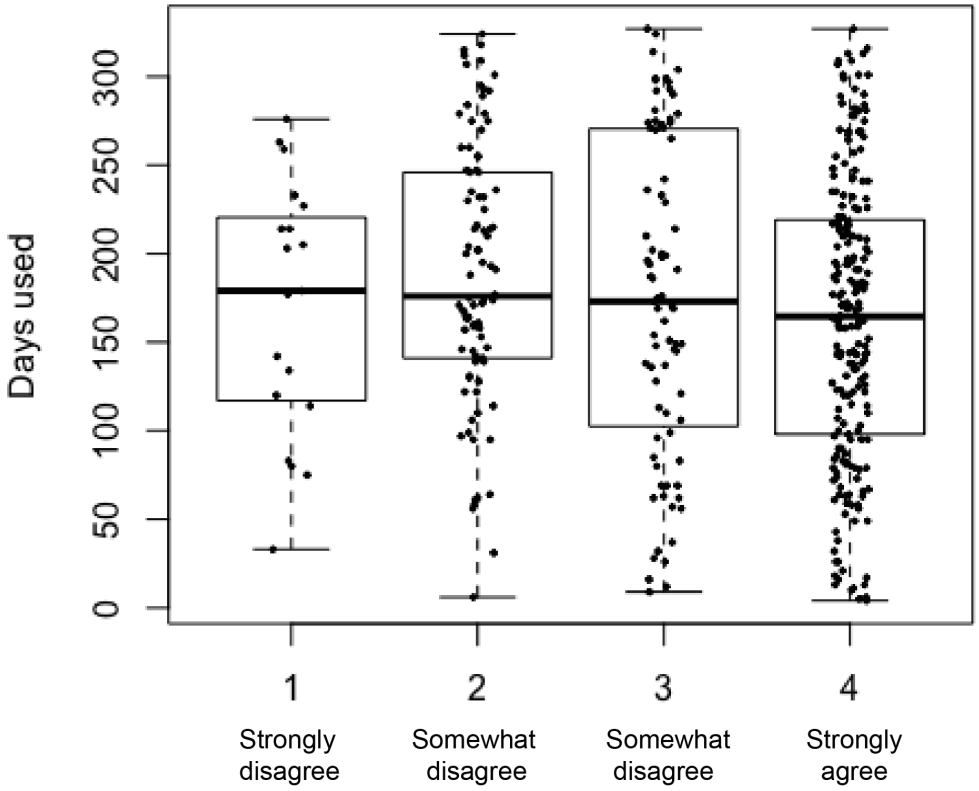


Figure 13. Plot of the partial dependence of sustained tracker use on the perceived effect of the tracker on goal attainment. A larger perceived effect leads to longer sustained use.

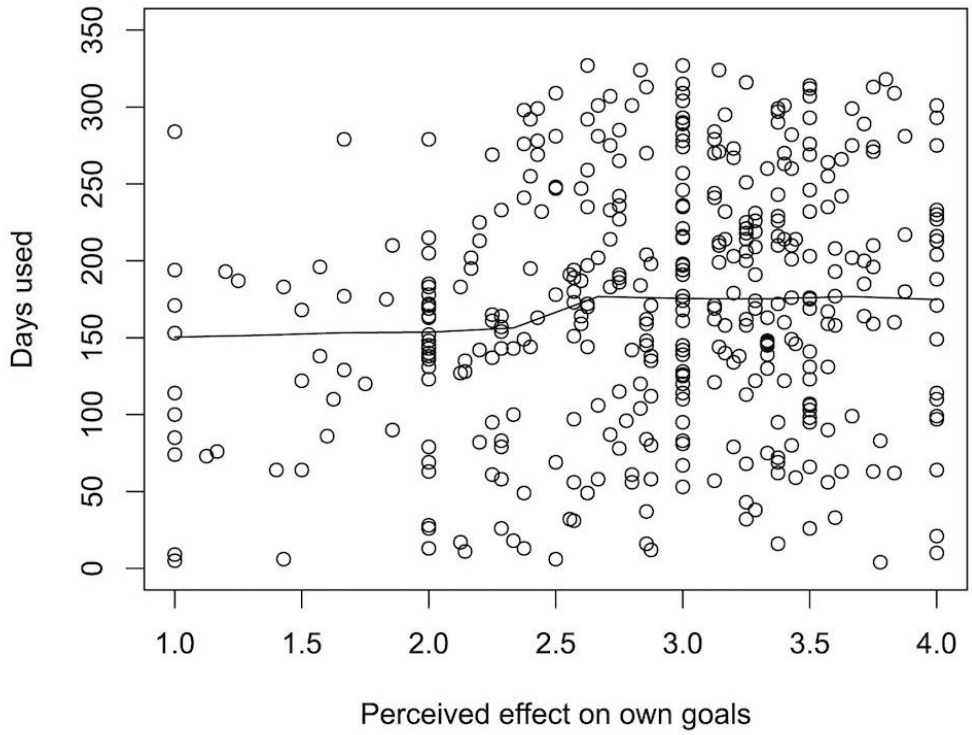


Figure 14. Plot of the partial dependence of sustained tracker use on user experience of the valence of the tracker. A better user experience leads to longer sustained use.

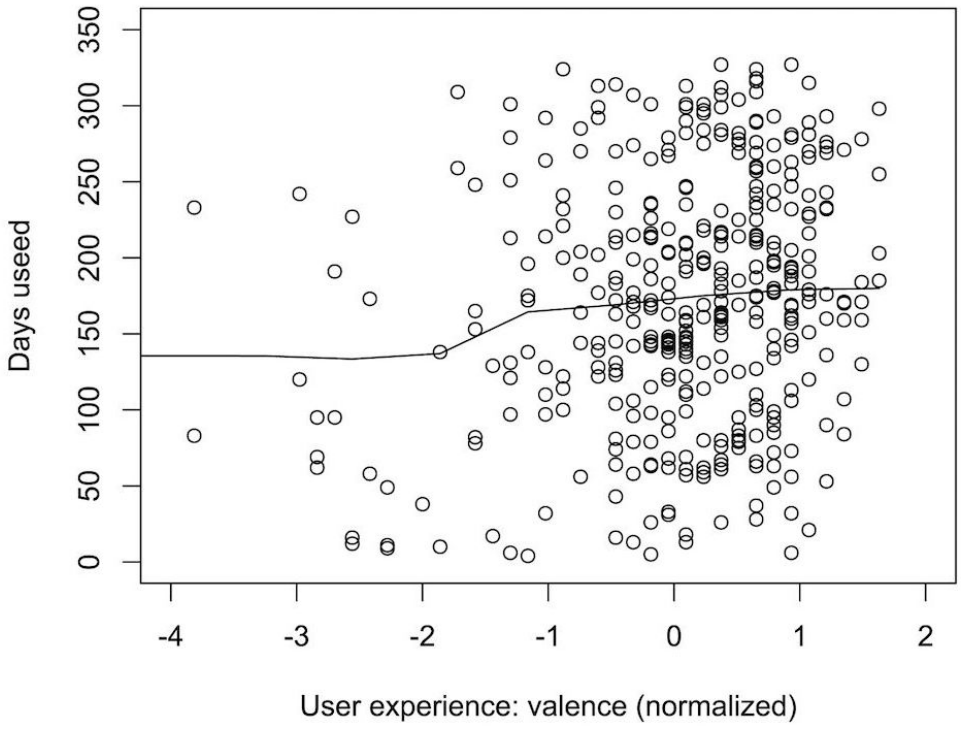
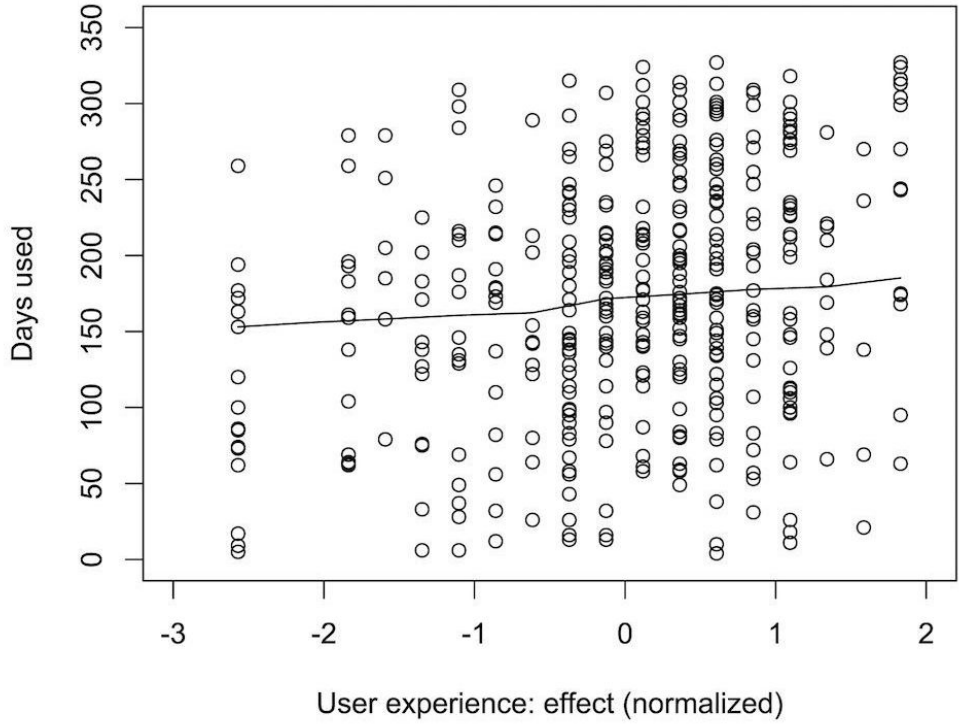


Figure 15. Plot of the partial dependence of sustained tracker use on user experience of the efficacy of the tracker. A better user experience leads to longer sustained use.



Discussion

Principal Findings

This study examined the use of an activity tracker, as well as reasons to stop using the tracker and predictors of sustained use. This study shows that of the 711 initial participants, approximately 50% still used their tracker after 6 months, and 12% continued to use their tracker even after 300 days. This rate of decay in usage confirms earlier findings [18]. This result also confirms the notion that wearable activity trackers are not subject to the rapid exponential decay of use we see in mobile phone apps, where usually 80% of users drop out in the first few days (eg, in [70]).

Reasons to Quit Tracking

After half a year, half of the participants quit using their tracker; at the sending out of Q3, three-quarters of participants were no longer using the Fitbit. When asked for reasons for their quitting, 56.7% of those who answered the question stated some form of technical malfunctioning (including empty batteries). A further 12.8% indicated they had lost the device. This result confirms findings (eg, [21]) in which trackers were abandoned because of empty batteries, with the perceived cost of replacement too high or too cumbersome [71], but this result deviates from other findings (eg, [22,24]) in which technological failures comprise only a very small part of reasons to no longer use a device, and reasons such as sustained motivation, device aesthetics, and device accuracy seem to be of more importance. However, in studies in which reasons to quit tracking are covered, either the demand characteristics of the study (eg, [24]) or the data gathering technique (eg, [22], where advertisement data from Craigslist were used) preclude the reliable registration of technical failure as a reason to quit tracking. The results from this study may therefore serve as a first indication of the relative importance of technological reliability for the sustained use of feedback technology. Simple actions, such as changing a battery, already seem to raise insurmountable barriers for sustained use of a tracker. Newer activity trackers, fortunately, mostly do not rely on button cell batteries, which take some effort to replace, but make it possible to recharge the device, much like one would recharge a smartphone. This, however, also constitutes a barrier to sustained use. Further research into the effect of technological failures on sustained use of activity trackers, and how to help users overcome the barriers brought about by technical issues such as empty barriers, is needed to corroborate this finding and shed light on potential solution strategies.

Relative Importance of Predictors for Self-Tracking

In the analysis of items from Q1 only (Figure 5), participant age was the only predictor that showed great impact; higher age was associated with more sustained use. A second group of predictors for sustained use were the goal to quit smoking (with not having the goal associated with longer use), iPhone type (with not having an iPhone associated with longer use), and household type (with single parents using the tracker for a shorter duration than all other groups). All other predictors were not found to have a noteworthy impact on sustained tracker use.

In the analysis of all (three) questionnaires (Figure 6), participant age was once again the strongest predictor of sustained tracker use. As in Q1, sustained use increased with age (to a point), overall (Q1 to Q3) the Fitbit was used for a shorter duration by the youngest age group (under 25). Other important predictors were user experience–related predictors (tracker valence and user experience, and perceived efficacy and helpfulness of the

tracker). iPhone type, having the goal to change eating habits, and wanting to quit smoking were other relevant predictors, albeit in an opposing way: these predictors were associated with a decreased use of the self-tracking device.

The relative importance of age as a predictor of sustained tracker use (with higher age associated with longer use) is in line with previous literature (eg, [25,30,31]). More research is needed to answer the question why this is, and to determine its implications for the design of tracker-based interventions for physical activity. Is the greater efficacy for older participants problematic, or are younger participants already well-served by other possibilities to exert themselves physically? In a similar notion, is this age-effect a consequence of self-selection, in which younger people already have enough alternatives for physical activity, and it is mostly those older than 25 years that turn to tracking as viable solution? Answers to these questions also have implications for the development of tracker interventions. Do we need more age-inclusive solutions, or can we regard this type of intervention as more effective for people older than 25 years?

The importance of user experience–related predictors such as valence, perceived efficacy, and preciseness of the tracker is also in line with previous studies. User experience and ease of use [22,23], functionality or lack thereof [22], the possibility to upgrade toward a newer device (ibidem), aesthetics and form [24], and perceived fit between device and self-image [23] have all been cited as reasons to either abandon the tracker or to keep using it. This sheds light on the relative importance of technological and design-related aspects of behavior change feedback technology aimed at greater physical activity. Even though there is a substantive literature on the subject, this remains an underexposed area in current health behavior change research. Clunky intervention designs carry the risk of being rejected by their participants and, more importantly, a lack of uptake once the intervention hits the market or the app store. In health behavior change research, a lot more attention is needed for user experience, user friendliness, and the aesthetic experience.

The negative impact of goals, such as wanting to change one's eating habits or wanting to quit smoking, seems logical in hindsight. The Fitbit tracker does not in itself contribute to the attainment of these goals, which could easily have a demoralizing effect. The fact that having an iPhone seemed to reduce the chances of sustained use could point at another covert measure of user experience. The iPhone interface for Fitbit-related feedback might possibly be more difficult to use or less functional than its Android equivalent; alternatively, iPhone users may be psychographically different from Android users on traits that lead to reduced usage of activity trackers. However, no evidence to support either hypothesis is presently available.

A surprising finding was the lack of effect of a range of predictors, which is not in line with previous literature (eg, [12,20]). *Socioeconomic status* markers such as education (eg, [25,27]) and profession [28,29], *gender* (eg, [25,27,32]), *psychological traits* (eg, [33-35]), *personal health-related factors* (eg, [36,37]), strong *motivation* (eg, in [30,31]), *strong, clear goals* (eg, in [25]), and *social interaction* (eg, in [50]) did not appear to affect tracker use. These have often been researched out of context, with predictors singled out and assessed independently. The current result could point to the fact that some predictors may not be as important as we think they are, when compared with many other possibilities. When placed in context, their role may be smaller than we assumed. A competing hypothesis, however, could be preselection; for instance, it is possible that motivation did not play a large role, because those who entered the challenge were already highly motivated. Similarly, perhaps only those already high on psychological traits such as conscientiousness took part. This preselection would limit the confidence in some of the null-results found in this study. If so, however, this preselection constitutes less of a problem as one would think. We can assume similar preselection would take place in the market place; it is reasonable to suspect that traits and states found in those who take part in this study would resemble states and traits of those people who would be interested in using a Fitbit in the first place. Unfortunately, this cannot be deducted from our research. Further research would be interesting.

The entire range of independent variables in Q1 explained 8.29% of variance; in Q1-Q3, the whole set of predictors accounted for 10.91% of explained variance. In Cohen's [72] frequently used assessments of effect sizes for psychology, an R^2 of .095 (Q1) to .099 (Q1-Q3) are described as a small effect or approaching a medium effect. Such an effect size is common in social and behavioral sciences, for situations where there is a lot of individual variation and many different factors may affect the dependent variable independently (see also [73]). To our best knowledge, this is the first quantitative study looking into the factors influencing the persistent use of activity trackers. Earlier studies (eg, [21,48,49,51-53]) were qualitative and small-scale studies (7 to 31 participants) and did not attempt to model activity tracker use. Thus, we have no immediate context to compare our model's performance with.

Intuitively, we may have expected a larger effect size from such a broad range of predictors. We can discern two competing hypotheses. A first hypothesis is that sustained use is mostly predicted by random events such as empty batteries or loss, but there are many small but significant contributions from a broad range of predictors. A second hypothesis is that unmeasured third variables are responsible for the relative lack of effect. Not all relevant predictors we could identify in the literature were included in the questionnaires. First, perceived self-efficacy was not directly assessed but only through measures regarding perceived efficacy of healthier behavior change. Second, literature [49] suggests that different tracking styles exist, such as tracking physical activity to diagnose a secondary problem such as sleeping disorders or stomach problems, or "fetishized" tracking: tracking because it is cool or otherwise desirable. In this study, the tacit assumption is that all participants want to at least document and probably also change their physical activity, which might not be the case in reality. Third, different forms of intrinsic motivation, such as motivation for autonomy, mastery, and relatedness [74], might lead to different levels of adherence to activity tracking. Finally, the completion of set goals was not registered. It is plausible to assume that when

people achieve their goal, their interest in tracking their progress wanes. The inclusion of these possible moderators in future research would shed light on their effect on sustained use of a tracking device. Further research could shed light on which of these possible explanations would be most feasible.

Limiting Factors

A few limitations to this study warrant further discussion. First, our confidence in the validity of the findings is limited by the fact that of the original 929 participants, 711 gave permission to access their Fitbit data and filled out Q1; of those 711, only 575 took part in Q2 (80.9%), and 542 took part in both Q2 and Q3 (76.2%). The greater part of those participants who did not fill out Q2 or Q3 quit using their Fitbit somewhere in the period preceding that questionnaire. Even though this decline in adherence is not at all uncommon in interventions for health behavior change, and thereby no cause for alarm, their data would have increased the validity and reliability of our findings.

Similarly, 56.7% of those who provided a reason for their no longer tracking stated technical malfunctioning. Of those who did not report a reason (eg, because they did not fill out Q2 or Q3), we do not know why they no longer took part. However, the fact that at least 17.5% of all participants quit because of technical reasons still emphasizes the importance of this finding, regardless of the reasons the nonreporters could have had for quitting.

A second, and possibly greater, limitation to the validity of the findings stems from the way the study design was carried out. Questionnaire construction and data gathering were carried out by the MySantéMobile team. The quality of the questionnaires would have benefited from early involvement of social scientists with relevant experience in questionnaire construction, which would have led to a more hypothesis-based selection of questions, and more informative response scales. As it is, we think this study has enough validity to serve its purpose, that is, as an exploratory analysis of potential determinants of sustained use of physical activity trackers.

A third limitation in the study design is the fact that the psychological predictors of use such as the Big Five and user experience–related predictors were not included in Q1 but made a first appearance in Q3. This limits the applicability of findings concerning these predictors because only participants making it to Q3 (76.2% of those who filled out Q1 and gave access to their data) answered these questions. However, psychological traits are known to be stable [75], so it is reasonable to expect that no great changes in big five traits occurred. User experience-related predictors can only be measured once participants have used the product; an a-priori judgment lacks value. These, therefore, could not have been included in Q1.

Finally, the data analysis method selected has its benefits, such as robustness toward overfitting and good handling of relatively low participant populations, but Random Forest analysis also has its limitations. The result of the analysis is a ranking of the relative importance of each predictor on the use of the activity tracker. Due to its ensemble nature, results from a Random Forest can be hard to interpret (unlike a linear model). Contributions from a variable can be present in multiple ways and through nonlinear and/or (higher-order) interactions. However, through Random Forest modeling, we can establish which predictor variables are important with respect to outcome variables. These variables can be studied further to establish their effect and interactions with other variables.

Conclusions

This study confirms earlier findings that habitual use of an activity tracker tends to decline at a slow exponential pace rather than show the rapid exponential decline shown in health app use. When they start using an activity tracker, most users in our sample continued to use it for at least half a year. Around 12% of users still use their tracker after 300 days.

This study also shows that sustained use of an activity tracker is not easy to predict. Most known predictors of sustained adherence to physical activity interventions do not seem to have an impact on sustained use in the sample observed in this study. When participants no longer use their tracker, technological failures such as empty batteries seem the predominant reason to quit.

The broad range of predictors entered in the Random Forest model in this study only led to a small proportion of explained variance. Those predictors that did have an effect on sustained use were participant age and factors related to the user experience of tracker use.

Regardless of the limitations to the findings cited above, this study shows some much-needed insight in predictors of sustained use of trackers. Furthermore, this study is one of few examples in which academia gets the chance to evaluate data from industry; the field would greatly benefit from a greater number of such collaborations, preferably with a larger role for the academic partner in setting up the study.

Acknowledgments

The authors would like to thank IDS Santé and everyone involved in the MySantéMobile project for their work in setting up the project and collecting the data and Jelmer Wolterink for his invaluable advice on machine learning techniques.

Conflicts of Interest

None declared.

Multimedia Appendix 1

All questionnaire items with response scales, variables in which they were used, transformations, and validity evaluation.

[PDF File (Adobe PDF File), 214KB - mhealth_v5i10e164_app1.pdf]

Multimedia Appendix 2

Principal Component Analysis of items relating to User Experience.

[PDF File (Adobe PDF File), 53KB - mhealth_v5i10e164_app2.pdf]

References

1. World Health Organization. apps.who.int. 2010. Global Database on Obesity URL: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html [accessed 2016-12-13] [WebCite Cache ID 6miv5y5AC]
2. Pedersen BK, Saltin B. Evidence for prescribing exercise as therapy in chronic disease. Scand J Med Sci Sports 2006 Feb;16(Suppl 1):3-63. [doi: 10.1111/j.1600-0838.2006.00520.x] [Medline: 16451303]
3. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. CMAJ 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
4. 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 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
5. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
6. Kremers SP, Brug J. Habit strength of physical activity and sedentary behavior among children and adolescents. Pediatr Exerc Sci 2008 Feb;20(1):5-14; discussion 14. [Medline: 18364530]
7. Phillips LA, Gardner B. Habitual exercise instigation (vs. execution) predicts healthy adults' exercise frequency. Health Psychol 2016 Jan;35(1):69-77. [doi: 10.1037/hea0000249] [Medline: 26148187]
8. Wilson TD. Strangers to Ourselves - Discovering the Adaptive Unconscious. Cambridge, Mass: Belknap Press of Harvard University Press; 2002.
9. Godino JG, Watkinson C, Corder K, Sutton S, Griffin SJ, van Sluijs EM. Awareness of physical activity in healthy middle-aged adults: a cross-sectional study of associations with sociodemographic, biological, behavioural, and psychological factors. BMC Public Health 2014 May 2;14:421 [FREE Full text] [doi: 10.1186/1471-2458-14-421] [Medline: 24886612]
10. 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 Oct 27;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]
11. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007 Mar 1;77(1):81-112. [doi: 10.3102/003465430298487]
12. Hermsen S, Frost JH, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Human Behav 2016;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
13. Verplanken B, Wood W. Interventions to break and create consumer habits. J Public Policy Mark 2006;25(1):90-103. [doi: 10.1509/jppm.25.1.90]
14. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
15. Pewinternet. 2013. Tracking for Health URL: http://www.pewinternet.org/2013/01/28/tracking-for-health/ [accessed 2016-12-15] [WebCite Cache ID 6mm2N2YSM]
16. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. J Am Med Assoc 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
17. H-Jennings F, Clément MV, Brown M, Leong B, Shen L, Dong C. Promote students' healthy behavior through sensor and game: a randomized controlled trial. Med Sci Educ 2016 Sep;26(3):349-355. [doi: 10.1007/s40670-016-0253-8]
18. Thorup C, Hansen J, Grønkjær M, Andreasen JJ, Nielsen G, Sørensen EE, et al. Cardiac patients' walking activity determined by a step counter in cardiac telerehabilitation: data from the intervention arm of a randomized controlled trial. J Med Internet Res 2016 Apr 4;18(4):e69 [FREE Full text] [doi: 10.2196/jmir.5191] [Medline: 27044310]
19. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, 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. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]
20. Fischer C. Feedback on household electricity consumption: a tool for saving energy? Energ Effic 2008;1:79-104. [doi: 10.1007/s12053-008-9009-7]
21. Harrison D, Berthouze N, Marshall P, Bird J. Tracking physical activity: problems related to running longitudinal studies with commercial devices. New York, NY, USA: ACM; 2014 Presented at: UbiComp '14 Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication; September 13-17; Seattle, Washington p. 699-702. [doi: 10.1145/2638728.2641320]

22. Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. New York, NY, USA: ACM; 2015 Presented at: UbiComp '15 Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11; Osaka, Japan p. 647-658. [doi: 10.1145/2750858.2807554]
23. Lazar A, Koehler C, Tanenbaum J, Nguyen DH. Why we use and abandon smart devices. New York, NY, USA: ACM; 2015 Presented at: UbiComp '15 Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11; Osaka, Japan p. 635-646. [doi: 10.1145/2750858.2804288]
24. Harrison D, Marshall P, Bianchi-Berthouze N, Bird J. Activity tracking: barriers, workarounds and customisation. New York, NY, USA: ACM; 2015 Presented at: UbiComp '15 Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11; Osaka, Japan p. 617-621. [doi: 10.1145/2750858.2805832]
25. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
26. Peels DA, Bolman C, Golsteijn RH, de Vries H, Mudde AN, van Stralen MM, et al. Differences in reach and attrition between web-based and print-delivered tailored interventions among adults over 50 years of age: clustered randomized trial. J Med Internet Res 2012;14(6):e179. [doi: 10.2196/jmir.2229]
27. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. Health Promot Int 2013 Sep;28(3):442-452 [FREE Full text] [doi: 10.1093/heapro/das029] [Medline: 22786673]
28. Al-Asadi AM, Klein B, Meyer D. Pretreatment attrition and formal withdrawal during treatment and their predictors: an exploratory study of the anxiety online data. J Med Internet Res 2014 Jun 17;16(6):e152 [FREE Full text] [doi: 10.2196/jmir.2989] [Medline: 24938311]
29. Habibović M, Cuijpers P, Alings M, van der Voort P, Theuns D, Bouwels L, et al. Attrition and adherence in a WEB-Based Distress Management Program for Implantable Cardioverter defibrillator Patients (WEBCARE): randomized controlled trial. J Med Internet Res 2014 Feb 28;16(2):e52 [FREE Full text] [doi: 10.2196/jmir.2809] [Medline: 24583632]
30. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
31. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
32. Wanner M, Martin-Diener E, Bauer G, Braun-Fahrländer C, Martin BW. Comparison of trial participants and open access users of a web-based physical activity intervention regarding adherence, attrition, and repeated participation. J Med Internet Res 2010 Feb 10;12(1):e3 [FREE Full text] [doi: 10.2196/jmir.1361] [Medline: 20147006]
33. Baumeister RF, Heatherton TF. Self-regulation failure: an overview. Psychol Inq 1996;7(1):1-15. [doi: 10.1207/s15327965pli0701_1]
34. Braverman J. Testimonials versus informational persuasive messages: the moderating effect of delivery mode and personal involvement. Communic Res 2008 Aug 4;35(5):666-694. [doi: 10.1177/0093650208321785]
35. Kuhl J. Action Control. Berlin, Heidelberg: Springer; 1985:267-275.
36. Grembowski D, Patrick D, Diehr P, Durham M, Beresford S, Kay E, et al. Self-efficacy and health behavior among older adults. J Health Soc Behav 1993 Jun;34(2):89-104. [Medline: 8277130]
37. Strecher VJ, DeVellis BM, Becker MH, Rosenstock IM. The role of self-efficacy in achieving health behavior change. Health Educ Q 1986;13(1):73-92. [Medline: 3957687]
38. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
39. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
40. 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]
41. Wilson K, Brookfield D. Effect of goal setting on motivation and adherence in a six‐week exercise program. Int J Sport Exerc Psychol 2009;7(1):89-100. [doi: 10.1080/1612197x.2009.9671894]
42. Wang J, Sereika SM, Chasens ER, Ewing LJ, Matthews JT, Burke LE. Effect of adherence to self-monitoring of diet and physical activity on weight loss in a technology-supported behavioral intervention. Patient Prefer Adherence 2012;6:221-226 [FREE Full text] [doi: 10.2147/PPA.S28889] [Medline: 22536058]
43. Murnane EL, Huffaker D, Kossinets G. Mobile health apps: adoption, adherence, and abandonment. New York, NY, USA: ACM; 2015 Presented at: UbiComp/ISWC'15 Adjunct Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers; September 7-11; Osaka, Japan p. 261-264. [doi: 10.1145/2800835.2800943]

44. Bandura A. Self-Efficacy. London, UK: Informa; 1992.
45. Tractinsky N, Katz AS, Ikar D. What is beautiful is usable. Interact Comput 2000;13(2):127-145. [doi: 10.1016/S0953-5438(00)00031-X]
46. Michie S, West R. A Guide to Development and Evaluation of Digital Interventions in Healthcare. London, UK: Silverback Publishing; 2016.
47. Canhoto AI, Arp S. Exploring the factors that support adoption and sustained use of health and fitness wearables. J Mark Manage 2017;33(1-2):32-60. [doi: 10.1080/0267257X.2016.1234505]
48. Fritz T, Huang EM, Murphy GC, Zimmermann T. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. New York, NY, USA: ACM; 2014 Presented at: CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Pages; April 26-May 01; Toronto, Ontario, Canada p. 487-496. [doi: 10.1145/2556288.2557383]
49. Rooksby J, Rost M, Morrison A, Chalmers MC. Personal tracking as lived informatics. New York, NY, USA: ACM; 2014 Presented at: CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26-May 01; Toronto, Ontario, Canada p. 1163-1172. [doi: 10.1145/2556288.2557039]
50. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
51. Munson SA, Consolvo S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. 2012 Presented at: 2012 6th International Conference on Pervasive Computing Technologies for Healthcare and Workshops, PervasiveHealth 2012; May 21-24; San Diego, CA p. 25-32. [doi: 10.4108/icst.pervasivehealth.2012.248691]
52. Rapp A, Cena F. Affordances for self-tracking wearable devices. New York, NY, USA: ACM; 2015 Presented at: ISWC '15 Proceedings of the 2015 ACM International Symposium on Wearable Computers; September 7-11; Osaka, Japan p. 141-142. [doi: 10.1145/2802083.2802090]
53. Shih PC, Han K, Poole ES, Rosson MB, Carroll JM. iConference 2015 Proceedings. 2015. Use and adoption challenges of wearable activity trackers URL: https://www.ideals.illinois.edu/bitstream/handle/2142/73649/164_ready.pdf?sequence=2&isAllowed=y
54. Gouveia R, Karapanos E, Hassenzahl M. How do we engage with activity trackers?: a longitudinal study of Habito. New York, NY, USA: ACM; 2015 Presented at: UbiComp '15 Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11; Osaka, Japan. [doi: 10.1145/2750858.2804290]
55. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]
56. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods 2009 Dec;14(4):323-348 [FREE Full text] [doi: 10.1037/a0016973] [Medline: 19968396]
57. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
58. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015 Oct 12;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
59. Bergkvist L, Rossiter JR. The predictive validity of multiple-item versus single-item measures of the same constructs. J Mark Res 2007 May;44(2):175-184. [doi: 10.1509/jmkr.44.2.175]
60. Cooke PJ, Melchert TP, Connor K. Measuring well-being: a review of instruments. Couns Psychol 2016;44(5):730-757. [doi: 10.1177/0011000016633507]
61. Jovanović V. The validity of the Satisfaction with Life Scale in adolescents and a comparison with single-item life satisfaction measures: a preliminary study. Qual Life Res 2016 Dec;25(12):3173-3180. [doi: 10.1007/s11136-016-1331-5] [Medline: 27262574]
62. Gosling SD, Rentfrow PJ, Swann Jr WB. PsycTESTS Dataset. 2003. Ten-Item Personality Inventory URL: http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Personality-TenItemPersonalityInventory.pdf [accessed 2017-10-17] [WebCite Cache ID 6uHHK4E2v]
63. International Standardization Organization (ISO). Ergonomics of human system interaction - Part 210. 2010. Human-centred design for interactive systems (formerly known as 13407) URL: https://www.iso.org/obp/ui/ [accessed 2017-10-17] [WebCite Cache ID 6uHG1yhW0]
64. Vermeeren AP, Law EL, Roto V, Obrist M, Hoonhout J, Väänänen-Vainio-Mattila K. User experience evaluation methods: current state and development needs. New York, NY, USA: ACM; 2010 Presented at: NordiCHI '10 Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries; October 16-20; Reykjavik, Iceland p. 521-530. [doi: 10.1145/1868914.1868973]
65. Winter D, Schrepp M, Thomaschewski J. Faktoren der User Experience. Systematische Übersicht über produktrelevante UX-Qualitätsaspekte. In: Mensch und Computer 2015. Berlin, Germany: De Gruyter Oldenbourg; 2015:33-41.
66. Hassenzahl M. The thing and I: understanding the relationship between user and product. In: Funology. Human-Computer Interaction Series. Dordrecht: Springer; 2003:31-42.

67. Walsh CG, Ribeiro JD, Franklin JC. Predicting risk of suicide attempts over time through machine learning. Clin Psychol Sci 2017;5(3):457-469. [doi: 10.1177/2167702617691560]
68. R-project. 2015. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2016-12-15] [WebCite Cache ID 6mlpYAVua]
69. Liaw A, Wiener M, Hägg U. Classification and Regression by randomForest. R News 2002.
70. Chen A. Andrewchen.co. 2016. New data shows losing 80% of mobile users is normal, and why the best apps do better URL: http://andrewchen.co/new-data-shows-why-losing-80-of-your-mobile-users-is-normal-and-that-the-best-apps-do-much-better/ [accessed 2016-05-24] [WebCite Cache ID 6hjwoLkyM]
71. Ledger D, McCaffrey D. Endeavourpartners. How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement URL: http://endeavourpartners.net/assets/Endeavour-Partners-Wearables-White-Paper-20141.pdf [accessed 2016-12-13] [WebCite Cache ID 6mjEFsA2b]
72. Cohen J. A power primer. Psychol Bull 1992;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155]
73. Abelson RP. A variance explanation paradox: when a little is a lot. Psychol Bull 1985;97(1):129-133. [doi: 10.1037/0033-2909.97.1.129]
74. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
75. Cobb-Clark DA, Schurer S. The stability of big-five personality traits. Econ Lett 2012;115(1):11-15. [doi: 10.2139/ssrn.1919414]

Abbreviations

BCT: behavior change technique
iOS: iPhone operating system
MSE: mean squared error
SES: socioeconomic status
SD: standard deviation

Edited by G Eysenbach; submitted 12.01.17; peer-reviewed by JN Kramer, T Kowatsch; comments to author 30.01.17; revised version received 06.07.17; accepted 31.07.17; published 30.10.17.

Please cite as:
Hermsen S, Moons J, Kerkhof P, Wiekens C, De Groot M
Determinants for Sustained Use of an Activity Tracker: Observational Study
JMIR Mhealth Uhealth 2017;5(10):e164
URL: http://mhealth.jmir.org/2017/10/e164/
doi:10.2196/mhealth.7311
PMID:29084709

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org